

Nadine Melloni Neumann

**Sistemas de Filas: Métodos de Simulação e
Métodos Analíticos**

Niterói - RJ, Brasil

11 de Dezembro de 2014

Nadine Melloni Neumann

**Sistemas de Filas: Métodos de
Simulação e Métodos Analíticos**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador: Prof. Dr. Valentin Sisko

Niterói - RJ, Brasil

11 de Dezembro de 2014

Nadine Melloni Neumann

**Sistemas de Filas: Métodos de Simulação e
Métodos Analíticos**

Monografia de Projeto Final de Graduação sob o título “*Sistemas de Filas: Métodos de Simulação e Métodos Analíticos*”, defendida por Nadine Melloni Neumann e aprovada em 11 de Dezembro de 2014, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Prof. Dr. Valentin Sisko

Orientador

Departamento de Estatística – UFF

Prof. Dr. Freddy Rolando Hernandez Romero

Departamento de Matemática Aplicada – UFF

Prof. Dr. Luis Guillermo Coca Velarde

Departamento de Estatística – UFF

Niterói, 11 de Dezembro de 2014

N492 Neumann, Nadine Melloni
Sistemas de filas: métodos de simulação e métodos analíticos
/ Nadine Melloni Neumann; Valentin Sisko, orientador.
Niterói, 2014.
69 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística) - Universidade Federal Fluminense,
Instituto de Matemática e Estatística, Niterói, 2014.

1. Teoria da fila. 2. Método de simulação. 3. Cadeia de
Markov. 4. Método analítico. I. Sisko, Valentin, orientador.
II. Universidade Federal Fluminense. Instituto de Matemática e
Estatística. III. Título.

CDD -519.82

Resumo

As filas são fenômenos que acontecem a todo o momento em nossos dias. Na teoria de filas, duas classes de métodos podem ser aplicadas para análise de problemas ligados aos sistemas de filas: métodos de simulação e métodos analíticos. O objetivo deste trabalho é comparar essas duas classes de métodos de análise. Serão comparados os resultados de algumas medidas de performances para três modelos, que serão apresentados ao longo deste trabalho, obtidas através da simulação via Cadeias de Markov e através do PDQ (Pretty Damn Quick), pacote disponível no software estatístico R que resolve analiticamente problemas ligados a sistemas de filas. Também será mostrado através de processos de nascimento e morte e Jackson Networks como o PDQ calcula essas medidas de performances.

Palavras-chaves: Filas, Simulação, Cadeias de Markov, PDQ, Processos de Nascimento e Morte, Jackson Networks.

*O que nos coloca em problemas
não é aquilo que não sabemos.
É o que sabemos com certeza,
e simplesmente está errado.*

Mark Twain

Agradecimentos

Agradeço e dedico este trabalho aos meus pais que não mediram esforços para que eu chegasse até esta etapa da minha vida. Minha mãe, Neusa, que com seu dom para psicóloga me ouviu e ajudou com belas palavras em todos os momentos que precisei e meu pai, Gilberto, com seu poder de acreditar em mim, muito mais que eu mesma, me deu sempre muita força e não deixou que eu desanimasse nos momentos difíceis. Agradeço também a toda minha família, pelo apoio, orações e cada palavra de carinho e incentivo.

Gostaria de agradecer a todos os meus professores, aos professores da UFF, sempre muito atenciosos, me instruíram e colaboraram com minha formação durante esses anos. À professora Jéssica, pelas valiosas sugestões e orientações para este trabalho e por estar sempre presente. Ao Valentin, que me orientou neste trabalho e me ajudou sempre que precisei durante toda a minha graduação.

Quero agradecer a todos que contribuíram na minha formação profissional. Ao Amauri, que acreditou em mim e sempre com muito humor me ajudou em minha primeira experiência profissional e a todos da NBS que me acolheram com muito carinho. Ao Jorge e a todos da Procuradoria que me mostraram e inseriram em um ambiente totalmente novo para mim. Ao Carlos Henrique por acreditar tanto em mim, à Maíra que me ajuda e compartilha conhecimentos todos os dias e a todos do IPEA que estão fazendo parte dessa experiência enriquecedora.

Agradeço de maneira geral à UFF que me proporcionou muitos momentos inesquecíveis e pessoas que vou levar para vida toda. Como o João, que de uma maneira admirável me apoia e me dá forças sempre que preciso e mesmo com toda essa distância. Como todos meus amigos da Estatística, Cissa e Luciana que são amigas 24 horas por dia, Guilherme, Natan, Pablo, Paola, Lucas, Dani, Bruno, Everson, Leo, Fábio e Thalita, que fizeram esses anos mais alegres e divertidos. Obrigada às meninas que dividem não só o apartamento mas o dia a dia comigo e a todos que partilharam preocupações, dúvidas, estudos e momentos de felicidades.

Finalizo agradecendo a Deus por ter me dado todas essas pessoas especiais, sei que sozinha isso não seria possível, pela minha saúde e a de todos que seguem ao meu lado. E a Nossa Senhora pela sua interseção e proteção.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Objetivos	p. 13
3	Metodologia	p. 14
3.1	Modelo de Filas	p. 14
3.1.1	Notação Kendall	p. 16
3.1.2	Medidas de Performance	p. 17
3.2	Processo Estocástico	p. 18
3.3	Distribuição Exponencial	p. 18
3.4	Processo de Poisson	p. 19
3.4.1	Processo de contagem	p. 20
3.4.2	Processo de Poisson Homogêneo	p. 20
3.4.3	Tempo entre eventos	p. 21
3.5	Cadeia de Markov	p. 23
3.5.1	Caso discreto	p. 23
3.5.1.1	Equações de Chapman-Kolmogorov	p. 23
3.5.1.2	Classificação de estados	p. 24
3.5.1.3	Ergodicidade e recorrência positiva	p. 29

3.5.2	Caso contínuo	p. 29
3.6	Processo de Nascimento e Morte	p. 30
3.6.1	Exemplo de um nó simples	p. 32
3.6.2	Exemplo de um nó com servidor duplo	p. 34
3.7	Redes de Jackson	p. 39
3.7.1	Exemplo de um nó simples	p. 41
3.7.2	Medidas de performance para Redes de Jackson	p. 41
4	Modelos Utilizados	p. 42
4.1	Modelo 1 - Dois servidores em paralelo e fila única	p. 42
4.1.1	Cadeias de Markov - Modelo 1	p. 43
4.1.2	Medidas de performance - Modelo 1	p. 44
4.2	Modelo 2 - Dois servidores em série	p. 45
4.2.1	Cadeias de Markov - Modelo 2	p. 46
4.2.2	Equações de trânsito - Modelo 2	p. 46
4.2.3	Medidas de performance - Modelo 2	p. 47
4.3	Modelo 3 - Rede de Comunicações	p. 49
4.3.1	Cadeias de Markov - Modelo 3	p. 50
4.3.2	Equações de trânsito - Modelo 3	p. 50
4.3.3	Medidas de performance - Modelo 3	p. 50
5	Simulação	p. 53
5.1	Simulação - Modelo 1	p. 53
5.2	Simulação - Modelo 2	p. 54
5.3	Simulação - Modelo 3	p. 60
6	Resultados	p. 63
6.1	Resultados - Modelo 1	p. 63

6.2	Resultados - Modelo 2	p. 64
6.3	Resultados - Modelo 3	p. 64
7	Conclusões	p. 66
	Referências	p. 67

Lista de Figuras

1	Diagrama de transições para a cadeia de Markov do Exemplo 1	p. 25
2	Diagrama de transições para a cadeia de Markov do Exemplo 2	p. 26
3	Diagrama de transições para a cadeia de Markov do Exemplo 3	p. 28
4	Diagrama de transições para a cadeia de Markov do Exemplo 4	p. 28
5	Representação do sistema com um nó	p. 32
6	Sistema de um nó com servidor duplo	p. 35
7	Sistema de fila única e dois servidores homogêneos	p. 43
8	Diagrama de transição de estados para o Modelo 1.	p. 43
9	Sistema de dois servidores heterogêneos em série	p. 46
10	Diagrama de transição de estados para o Modelo 2.	p. 47
11	Sistema de uma rede de comunicações	p. 49

Lista de Tabelas

1	Resultados do Modelo 1	p. 63
2	Resultados do Modelo 2	p. 64
3	Resultados do Modelo 3	p. 65

1 Introdução

As filas são fenômenos que acontecem a todo o momento em nossos dias. Podemos ver fila em tudo o que fazemos e aonde vamos. Temos filas de carros, para comprar um lanche ou quando vamos pagar nossa conta. Em diversas e variadas situações podemos vê-las. As filas que acontecem no nosso dia-a-dia não são apenas as visíveis diretamente, existem filas nos aeroportos e portos e estão presentes em todo o processo de produção e escoamento de produtos, fazendo com que os preços das mercadorias subam ou desçam conforme a eficiência na distribuição e na logística. Desta forma, o estudo dos modelos de filas tem como objetivo a melhoria de desempenho do sistema, entendida, entre outros aspectos, como melhor utilização dos recursos de serviço disponíveis, menor tempo de espera, mais rapidez no atendimento e redução de custos.

Na Teoria de filas, duas classes de métodos podem ser aplicadas para a análise de problemas ligados aos sistemas de filas: métodos de simulação e métodos analíticos. Ambos os métodos têm suas qualidades e restrições.

A princípio, a simulação permite que sejam incluídas características específicas, como tempo de chegada ou atendimento não-homogêneos e como mais de uma classe de clientes (clientes que chegam à uma loteria para pagar contas ou fazer apostas). Porém a simulação é um método mais demorado, tanto na preparação quanto na execução.

Já para o método analítico será utilizado o PDQ (*Pretty Damn Quick*), pacote disponível no software estatístico R que resolve modelos de filas analiticamente. Este é um método extremamente rápido, basta saber passar o modelo através das funções. Mas apresenta restrições para os modelos que podem ser resolvidos, por exemplo, os tempos entre chegadas e de atendimentos devem ser exponenciais e, neste método também não é possível calcular estatísticas de ordem superior, como a variância das medidas de performance, o que é possível através da simulação. Neste trabalho só serão abordados modelos que podem ser analisados pelos dois métodos.

O PDQ é uma biblioteca de funções para a solução de modelos de desempenho, ex-

pressa como circuitos ou redes de filas associadas ao livro [1] e ao [2], e é fornecido como código aberto, escrito na linguagem C, para permitir que ele seja usado no ambiente de sua escolha. Existem interfaces para linguagens de programação C, Perl, Python e para o R que foi o utilizado neste trabalho. Esta técnica de solução algorítmica é mais rápida do que a simulação e outras técnicas. A desvantagem é não poder computar estatísticas de ordem superior, por exemplo a variância.

O Capítulo 2 apresenta os objetivos deste trabalho. Já no Capítulo 3, são apresentadas definições de Cadeia de Markov em tempo discreto e contínuo, definições de quando elas são irredutíveis e quando são ergódicas, definições de Redes de Jackson e propriedades delas e uma introdução de teoria de filas seguindo os Capítulos 4,5 e 6 do livro [3] e o Capítulo 4 do livro [4].

Os três modelos utilizados nas análises serão apresentados no Capítulo 4. Nesse mesmo capítulo, para cada modelo, são apresentadas as cadeias de Markov, equações de trânsito (em alguns casos), o cálculo analítico das medidas de performance e como é passado para o PDQ.

Os algoritmos utilizados na simulação desses modelos serão apresentados no Capítulo 5. Já os resultados encontrados pelo PDQ e pela simulação serão apresentados e analisados no Capítulo 6. As conclusões serão apresentadas no Capítulo 7.

2 Objetivos

O objetivo deste trabalho é comparar as duas classes de métodos de análise: método de simulação e método analítico. Serão comparadas as medidas de performance para três modelos que serão apresentados ao longo deste trabalho obtidas através da simulação via Cadeias de Markov e através do método analítico.

Também será mostrado através de processos de nascimento e morte e Redes de Jackson como o PDQ calcula as medidas de performance. Para isso, também foram estudados princípios de Cadeia de Markov em tempo discreto e contínuo.

3 Metodologia

Neste capítulo será apresentada a teoria para a construção deste trabalho. Contém explicações, definições e pequenos exemplos sobre Modelo de Filas, Cadeias de Markov e sobre o PDQ.

3.1 Modelo de Filas

Um sistema simples de filas consiste na chegada de clientes em um servidor para serem atendidos. Os clientes são quaisquer entidades que solicitem serviços do sistema. Os servidores são os recursos do sistema que fornecem o atendimento aos clientes. Uma fila é formada quando o servidor não possui disponibilidade imediata e alguns clientes precisam aguardar para serem atendidos.

O termo *cliente* é genérico em teoria das filas, aplicando-se não somente a seres humanos. O conceito pode representar:

- processos esperando para receber a CPU (Unidade Central de Processamento);
- pessoas em um caixa de supermercado;
- pacotes que chegam a um roteador para serem encaminhados;
- barcos esperando para atracarem no porto;
- equipamentos para serem concertados.

Desse mesmo modo, o *servidor* pode representar:

- CPU;
- caixa de supermercado;
- roteador;

- local onde os barcos atracam no porto;
- profissional que concerta equipamentos em uma assistência técnica.

Os modelos de filas são excelentes ferramentas para projetos de novos sistemas de filas e também para identificação de oportunidades de melhorias em sistemas existentes, fornecendo medidas de desempenho como percentual de utilização do servidor, comprimento das filas ou tempo de espera dos clientes. Estas medidas podem ser estimadas em função dos parâmetros de entrada do sistema, como a taxa de chegada dos clientes, o tempo de atendimento e o número e configuração dos servidores. Uma vez que alguns desses parâmetros de entrada estão sobre controle da administração do sistema, um modelo de fila ajuda a identificar a maneira mais adequada de utilizar os recursos existentes de forma a alcançar o melhor desempenho possível. Modelos relativamente simples podem ter solução matemática utilizando métodos analíticos de Teoria de Filas, o que representa um esforço significativamente menor do que com a utilização de simulação.

Conforme [5], um sistema de filas é caracterizado por seis componentes. Os três primeiros são obrigatórios e os três últimos, se não informados, são considerados conhecidos:

- Modelo de chegada dos usuários ao serviço: Sistemática que indica o intervalo entre as chegadas subsequentes dos clientes no sistema. As chegadas podem ser em instantes determinísticos ou aleatórios, assim como os clientes podem chegar um por vez, ou em grupos de tamanho constante ou aleatório.
- Modelo de serviço (atendimento aos usuários): Intervalo de tempo entre o início e o término da prestação de serviço ao cliente. Pode ser constante, variar aleatoriamente ou em função de outras variáveis, como tipo de cliente, horário ou comprimento da fila.
- Número de servidores (mecanismo de atendimento): Número e configuração dos servidores. Os sistemas de filas podem possuir um, múltiplos ou até mesmo infinitos servidores. Estes podem estar em sequência, paralelo ou em outra formação.
- Capacidade do sistema: é o número total de usuários que o sistema é capaz de atender. Esta capacidade também pode ser considerada finita ou infinita. Se este parâmetro não for informado, o sistema é considerado com capacidade ilimitada.
- Tamanho da população: Conjunto de clientes em potencial do sistema. A população pode ser finita ou infinita. Se este parâmetro não for informado, o tamanho da população é considerado infinito.

- Disciplina da Fila: Ordem lógica que determina qual é a sequência de atendimento dos clientes em fila. As mais comuns são FIFO (*First In, First Out*) quando o primeiro cliente que chega é o primeiro a ser atendido, LIFO (*Last In, First Out*) quando o último que chega é o primeiro a ser atendido e por prioridade de atendimento. Se a disciplina da fila não for informada, é considerada FIFO.

3.1.1 Notação Kendall

É a notação utilizada para descrever os modelos de filas. É necessário conhecer os seis componentes anteriores para especificar um sistema de filas. Esta notação recebe o nome de Notação de Kendall em homenagem ao seu criador David Kendall.

Sequência: $v/w/x/y/n/z$ onde:

v indica o modelo de chegada, alguns valores de v mais comuns são M(denota distribuição exponencial), G(distribuição geral), D(representa um tempo fixo, determinístico);

w denota o modelo do serviço, os valores são os mesmos de v ;

x significa o número de servidores;

y representa a capacidade do sistema, é omitido quando for infinito;

n representa o tamanho da população, também omitido quando for infinito e

z designa a disciplina da fila, quando for modelo FIFO, é omitido.

A notação condensada $A/B/c$ é muito usada e se assume que não há limite para o tamanho da fila, a população é infinita e a disciplina da fila é FIFO. Para A e B , quando a distribuição for exponencial, usa-se M (Markoviana).

O modelo M/M/1 diz respeito ao modelo de filas onde tanto as chegadas quanto o atendimento são markovianas (é o mesmo que dizer que seguem a distribuição Exponencial) e que temos um único servidor.

Existem vários outros modelos de fila:

- O modelo M/M/1/K (População Finita): um caso particular e bastante encontrado na vida prática é aquele onde a população de clientes é finita, tem tamanho K .

- O modelo M/M//C (População Infinita): aqui temos o modelo de uma única fila e C servidores e tanto chegada como atendimento são markovianas. Supõe-se aqui que a capacidade de atendimento de cada um dos atendentes é a mesma.
- O modelo M/M/C/K (População Finita): neste modelo encontramos C servidores e a população de tamanho K.

3.1.2 Medidas de Performance

Através da simulação é possível obter diversas medidas de performance se forem considerados alguns complicadores a fim de tornar a modelagem o mais fiel possível à realidade. Porém, neste trabalho, as medidas de performance serão limitadas às que são fornecidas pelo Pacote PDQ, uma vez que o objetivo é comparar os resultados.

Utilização

A utilização do sistema será apresentada em forma de porcentagem, ou seja, quanto do tempo total de funcionamento do sistema os servidores ficaram verdadeiramente ocupados.

Tamanho da fila

O tamanho da fila será a média, em relação ao tempo, do tamanho das filas em cada instante. Vale ressaltar que o cliente que está no servidor sendo atendido também pertence à fila.

Tamanho da fila de espera

Diferentemente do tamanho da fila, aqui o cliente que está em atendimento no servidor não é considerado. Como o próprio nome diz, o tamanho da fila de espera é a média dos tamanhos das filas de espera, em relação ao tempo.

Tempo de espera

O tempo de espera para cada cliente é a diferença do momento em que ele terminou de ser atendido, ou seja, o momento que ele saiu do sistema, pelo momento de chegada ao sistema. Se algum servidor estiver vazio no momento que um cliente chegar ao sistema este não entrará na fila, pois irá ser atendido imediatamente, logo o seu tempo de espera é zero.

Calculado o tempo de espera para cada cliente que passou pelo sistema, é feita uma média desses valores que nos fornecerá a medida de performance tempo de espera.

Tempo de permanência no sistema

A permanência no sistema para cada cliente é a diferença do momento de saída do sistema, quando terminou de ser atendido, pelo momento que ele chegou ao sistema (entrou na fila de espera ou foi direto ser atendido).

Dessa maneira, o tempo de permanência no sistema é a média desse tempo para cada indivíduo que passou pelo sistema.

3.2 Processo Estocástico

Segundo [6], um processo estocástico é uma família $(X_t)_{t \in \mathbb{T}}$ de variáveis aleatórias, onde o índice t representa o tempo. Informalmente falando, um processo estocástico descreve uma história que se desenvolve de uma forma aleatória ao longo de um período de tempo representado por \mathbb{T} . O \mathbb{T} pode ser igual ao conjunto de números naturais $\mathbb{N} = \{0, 1, \dots\}$, ao conjunto de números inteiros \mathbb{Z} , ao conjunto de números reais positivos $[0, +\infty)$ ou ao conjunto de números reais \mathbb{R} . Nos dois primeiros casos o processo será dito em tempo discreto e nos dois últimos casos ele será dito em tempo contínuo. Nos casos de processos em tempo discreto há uma preferência pela letra n para indicar o instante de tempo, deixando a letra t para os casos de processos em tempo contínuo. Suponhamos que as variáveis X_n ou X_t assumam valores no conjunto E . Esse conjunto será chamado de espaço de estados.

3.3 Distribuição Exponencial

A distribuição Exponencial é muito útil para descrever o tempo que se leva para completar uma tarefa ou tempo de duração de um equipamento, por exemplo. Algumas aplicações da distribuição exponencial incluem modelar o tempo entre ocorrências de eventos em um experimento de Poisson, desta forma, compreender tal distribuição e suas características é fundamental para o desenvolvimento deste trabalho.

Definição 3.3.1. *A variável contínua X tem distribuição Exponencial com parâmetro λ , $\lambda > 0$, se a função de densidade de probabilidade é*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{se } x < 0. \end{cases} \quad (3.1)$$

Logo, seja $X \sim \text{Exponencial}(\lambda)$, então sua função de distribuição acumulada será dada por:

$$F(x) = P\{X \leq x\} = \int_{-\infty}^x f(y)dy = \begin{cases} 1 - e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{se } x < 0. \end{cases} \quad (3.2)$$

E temos que o valor esperado e a variância de X são dados por:

$$E(X) = \int_0^{\infty} \lambda x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \quad (3.3)$$

$$E(X^2) = \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx = -x^2 e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda} \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{2}{\lambda^2} \quad (3.4)$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}. \quad (3.5)$$

Definição 3.3.2. *Uma variável aleatória X é sem memória, se*

$$P\{X > s + t \mid X > t\} = P\{X > s\} \quad \text{para todos } s, t \geq 0. \quad (3.6)$$

Interpretando X como o tempo da vida de algum instrumento, (3.6) afirma que a probabilidade que o instrumento viva $s + t$ horas dado que ele já viveu t horas é a mesma que a probabilidade inicial de ele viver s horas. Em outras palavras, se o instrumento é vivo no momento de tempo t , então a distribuição do tempo que ainda sobra para ele viver é a mesma que a probabilidade inicial do tempo da vida, isto é, o instrumento não lembra que ele já foi usado durante tempo t .

Observação 1. Outros dois jeitos de escrever (3.6) são

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\},$$

$$P\{X > s + t\} = P\{X > s\} P\{X > t\}. \quad (3.7)$$

3.4 Processo de Poisson

Um processo de Poisson pode ser dito como um caso particular de um processo de contagem [3] e por isto é fundamental entender o que é um processo de contagem para depois sim definirmos um processo de Poisson.

3.4.1 Processo de contagem

Definição 3.4.1. Um processo estocástico $(N(t))_{t \in [0, +\infty)}$ é um processo de contagem, se $N(t)$ representa o número de ‘eventos’ que ocorreram até o instante de tempo t .

Observação 2. Da Definição 3.4.1 imediatamente seguem as seguintes propriedades de um processo de contagem $(N(t))_{t \in [0, +\infty)}$:

- i. O espaço de estados é $E = \mathbb{N}$.
- ii. Se $s \leq t$, então $N(s) \leq N(t)$.
- iii. Se $s \leq t$, então $N(t) - N(s)$ é igual ao número de eventos que ocorreram no intervalo $(s, t]$.

Definição 3.4.2. Um processo de contagem tem incrementos independentes se os números dos eventos que ocorrem nos intervalos de tempos disjuntos são independentes.

Em outras palavras, se um processo de contagem possui incrementos independentes, o número de eventos que ocorrem até o instante s , $N(s)$, deve ser independente do número de eventos que ocorrem entre os instantes t e s , $N(t) - N(s)$.

Definição 3.4.3. Um processo de contagem tem incrementos estacionários se $N(t_2 + s) - N(t_1 + s)$ tem a mesma distribuição que $N(t_2) - N(t_1)$ para todos $t_1 < t_2$ e $s > 0$.

3.4.2 Processo de Poisson Homogêneo

Definição 3.4.4. Um processo de contagem $(N(t))_{t \in [0, +\infty)}$ é um processo de Poisson com a taxa λ , $\lambda > 0$, se

- i. $N(0) = 0$;
- ii. o processo tem incrementos independentes;
- iii. para quaisquer $s, t \geq 0$, temos

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \dots$$

Observação 3. Do item iii da Definição 3.4.4 segue que o processo de Poisson tem incrementos estacionários.

Observação 4. Para processo de Poisson, $E(N(t)) = \lambda t$. Isso explica a razão para chamar λ de taxa de processo de Poisson.

Definição 3.4.5. Um processo de contagem $(N(t))_{t \in [0, +\infty)}$ é um processo de Poisson com a taxa $\lambda, \lambda > 0$, se

- i. $N(0) = 0$;
- ii. o processo tem incrementos independentes;
- iii. $P\{N(h+t) - N(t) = 1\} = \lambda h + o(h), t \in [0, +\infty)$;
- iv. $P\{N(h+t) - N(t) \geq 2\} = o(h), t \in [0, +\infty)$.

3.4.3 Tempo entre eventos

Definição 3.4.6. Consideramos um processo de Poisson $(N(t))_{t \in [0, +\infty)}$. Denotamos o instante de ocorrência do primeiro evento por T_1 . Para $n \in \mathbb{N}^*$, denotamos T_n o comprimento do intervalo entre os instantes de ocorrência do $(n-1)$ -ésimo evento e n -ésimo evento. A sequência $(T_n)_{n \in \mathbb{N}^*}$ é chamada de sequência de tempos entre chegadas.

Se $T_1 = 5$ e $T_2 = 10$, por exemplo, então o instante do primeiro evento é 5 e do segundo é 15.

Proposição 3.4.7. $(T_n)_{n \in \mathbb{N}^*}$ são independentes e identicamente distribuídos com distribuição exponencial com o parâmetro λ .

Demonstração. Provamos primeiro que o instante T_1 tem a distribuição exponencial com parâmetro λ . Para isso notamos, que o evento $\{T_1 > t\}$ ocorre se e somente se nenhum evento do processo de Poisson ocorre durante o intervalo $[0, t]$, então

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}.$$

O que prova que T_1 tem a distribuição exponencial com o parâmetro λ . Agora,

$$P\{T_2 > t\} = E[P\{T_2 > t \mid T_1\}]. \quad (3.8)$$

Mas

$$\begin{aligned} P\{T_2 > t \mid T_1 = s\} &= P\{0 \text{ eventos ocorreram durante tempo } (s, s+t] \mid T_1 = s\} \\ &= P\{0 \text{ eventos ocorreram durante tempo } (s, s+t]\} = e^{-\lambda t} \end{aligned} \quad (3.9)$$

onde as últimas duas igualdades são válidas pois o processo de Poisson tem incrementos independentes e estacionários. Usando (3.9), podemos continuar o cálculo de (3.8):

$$\begin{aligned} P\{T_2 > t\} &= E[P\{T_2 > t \mid T_1\}] = \int_0^\infty P\{T_2 > t \mid T_1 = s\} \lambda e^{-\lambda s} ds \\ &= \int_0^\infty e^{-\lambda t} \lambda e^{-\lambda s} ds = e^{-\lambda t} \int_0^\infty \lambda e^{-\lambda s} ds = e^{-\lambda t} \end{aligned} \quad (3.10)$$

De (3.9) e (3.10) obtemos que T_1 e T_2 são independentes. Repetindo o raciocínio provamos a proposição. \square

A Proposição 3.4.7 não deve ser surpresa pois o fato que o processo possui os incrementos independentes faz o processo com cada incremento “esquecer” o passado, em outras palavras o processo possui a propriedade de ser “sem memória”.

Definição 3.4.8. *O instante quando ocorre o n -ésimo evento é denotado por S_n . S_n também é chamado de tempo de espera do n -ésimo evento.*

Proposição 3.4.9. *S_n tem distribuição Gama com parâmetros n e λ .*

Demonstração. Temos $S_n = T_1 + T_2 + \dots + T_n$. A proposição segue imediatamente da Proposição 3.4.7. Podemos fazer a demonstração utilizando um outro raciocínio. Temos

$$F_{S_n}(t) = P\{S_n \leq t\} = P\{N(t) \geq n\} = \sum_{k=n}^{\infty} P\{N(t) = k\} = \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Logo

$$\begin{aligned} f_{S_n}(t) &= (F_{S_n}(t))' = \sum_{k=n}^{\infty} \left(e^{-\lambda t} \frac{(\lambda t)^k}{k!} \right)' = \sum_{k=n}^{\infty} \left(-\lambda e^{-\lambda t} \frac{(\lambda t)^k}{k!} + \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} \right) \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} + \sum_{k=n+1}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} - \sum_{k=n}^{\infty} \lambda e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}. \quad \square \end{aligned}$$

A Proposição 3.4.7 oferece mais um jeito para definir o processo de Poisson.

Definição 3.4.10. *Seja $(T_n)_{n \in \mathbb{N}^*}$ uma seqüência de variáveis aleatórias independentes identicamente distribuídas com distribuição exponencial com parâmetro λ . Sejam $S_0 = 0$ e $S_n = T_n + S_{n-1}$, $n \in \mathbb{N}^*$. Definimos $N(t) = \max\{n : S_n \leq t\}$. $(N(t)) \in [0, +\infty)$ é processo de Poisson.*

3.5 Cadeia de Markov

3.5.1 Caso discreto

As duas seguintes definições podem ser encontradas em [3] e [6] correspondentemente. Além disso em [6] é demonstrado que elas são equivalentes.

Definição 3.5.1. *Um processo estocástico $(X_n)_{n \in \mathbb{N}}$, tendo como espaço de estados um conjunto enumerável E , é dito uma cadeia de Markov com matriz de transição $\mathbf{P} = (p_{xy})_{x,y \in E}$ se para todos os estados $x_0, x_1, \dots, x_{n-1}, x, y$ e todos $n \geq 0$, temos*

$$P\{X_{n+1} = y | X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = x_0\} = P\{X_{n+1} = y | X_n = x\} = p_{xy}. \quad (3.11)$$

Definição 3.5.2. *Um processo estocástico $(X_n)_{n \in \mathbb{N}}$, tendo como espaço de estados um conjunto enumerável E , é dito uma cadeia de Markov com função F se $F : E \times [0, 1] \rightarrow E$ e para todo $n \geq 1$ temos que*

$$X_n = F(X_{n-1}, U_n),$$

onde U_1, U_2, \dots é uma sequência de variáveis aleatórias independentes e distribuídas uniformemente no intervalo $[0, 1]$.

3.5.1.1 Equações de Chapman-Kolmogorov

A matriz $\mathbf{P} = (p_{ij})_{i,j \in E}$ representa as probabilidades de transição em um passo. A probabilidade de transição de i para j em n passos é

$$p_{ij}^{(n)} = P\{X_{n+m} = j | X_m = i\}, \quad n \in \mathbb{N}, \quad i, j \in E. \quad (3.12)$$

Evidentemente $p_{ij}^{(1)} = p_{ij}$. As equações de Chapman-Kolmogorov fornecem um método para calcular as probabilidades de transição em n -passos. Elas são

$$p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)}, \quad n, m \in \mathbb{N}, \quad i, j \in E. \quad (3.13)$$

e para entender que elas são válidas podemos notar que $p_{ik}^{(n)} p_{kj}^{(m)}$ representa a probabilidade que o processo começando no estado i vai para estado j em $n + m$ passos utilizando caminhos que passam pelo estado k no momento n . Então somando todos estados intermediários k obtemos a probabilidade de o processo estar no estado j no momento $n + m$.

Formalmente, temos

$$\begin{aligned} p_{ij}^{(n+m)} &= P\{X_{n+m} = j | X_0 = i\} = \sum_{k \in E} P\{X_{n+m} = j, X_n = k | X_0 = i\} \\ &= \sum_{k \in E} P\{X_{n+m} = j | X_n = k, X_0 = i\} P\{X_n = k | X_0 = i\} = \sum_{k \in E} p_{kj}^{(m)} p_{ik}^{(n)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)}. \end{aligned}$$

A matriz de transição em n passos é $\mathbf{P}^{(n)} = (p_{ij}^{(n)})_{i,j \in E}$. Evidentemente, $\mathbf{P}^{(1)} = \mathbf{P}$. Utilizando matrizes, a representação das equações (3.13) é

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)}.$$

Então, em particular, temos

$$\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P}^{(1)} \mathbf{P}^{(1)} = \mathbf{P} \mathbf{P} = \mathbf{P}^2$$

e por indução

$$\mathbf{P}^{(n)} = \mathbf{P}^{((n-1)+1)} = \mathbf{P}^{(n-1)} \mathbf{P}^{(1)} = \mathbf{P}^{n-1} \mathbf{P} = \mathbf{P}^n.$$

3.5.1.2 Classificação de estados

Definição 3.5.3. Um estado j chama-se acessível de um estado i se existe $n \in \mathbb{N}$ tal que $p_{ij}^{(n)} > 0$.

Observação 5. Pela definição, n pode ser igual à 0. Mas $p_{ij}^{(0)} = \begin{cases} 1 & \text{se } i = j, \\ 0 & \text{se } i \neq j. \end{cases}$

Isso significa que qualquer estado i é acessível dele mesmo.

Um estado j é acessível de um estado i se e somente se, começando no estado i , é possível que o processo entre no estado j .

Definição 3.5.4. Dois estados i e j se comunicam ($i \leftrightarrow j$), se i é acessível de j , e j é acessível de i .

Observação 6. Pela definição, qualquer estado i comunica com ele mesmo:

$$p_{ii}^{(0)} = P\{X_0 = i | X_0 = i\} = 1.$$

Proposição 3.5.5. Para a relação de comunicação as seguintes propriedades são válidas:

- i.* Qualquer estado i se comunica com ele mesmo, isto é, $i \leftrightarrow i$.

ii. Se um estado i se comunica com um estado j , então o estado j comunica com o estado i , isto é, $i \leftrightarrow j \Rightarrow j \leftrightarrow i$.

iii. Se um estado i se comunica com um estado j , e o estado j se comunica com um estado k , então o estado i se comunica com o estado k , isto é, $i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k$.

Demonstração. A propriedade (i) segue da Observação 6 e a propriedade (ii) da Definição 3.5.4. Para demonstrar a propriedade (iii) podemos usar as equações de Chapman-Kolmogorov (3.13). Vamos supor que i comunica com j , e j comunica com k . Pela Definição 3.5.3, existem $n \geq 0$ e $m \geq 0$ tais que $p_{ij}^{(n)} > 0$ e $p_{jk}^{(m)} > 0$. Temos

$$p_{ik}^{(n+m)} \stackrel{(3.13)}{=} \sum_{r \in E} p_{ir}^{(n)} p_{rk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0.$$

Logo o estado k é acessível do estado i . Do mesmo jeito podemos provar que o estado i é acessível do estado k . Então os estados i e k se comunicam. \square

Definição 3.5.6. Cadeia de Markov $(X_n)_{n \in \mathbb{N}}$ chama-se irredutível se todos os estados se comunicam entre si.

Exemplo 1. Consideramos uma cadeia de Markov com o espaço de estados $\{0, 1, 2\}$ e a matriz de transição

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

determinaremos se a cadeia de Markov com esta matriz de transição é irredutível.

Solução. Temos

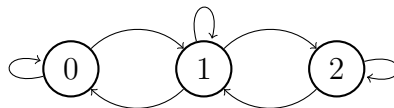


Figura 1: Diagrama de transições para a cadeia de Markov do Exemplo 1

Na Figura 1 estamos vendo que $0 \leftrightarrow 1$ e $1 \leftrightarrow 2$. Pela propriedade (iii), obtemos $0 \leftrightarrow 2$. Então todos os estados são comunicáveis. \square

Exemplo 2. Consideramos uma cadeia de Markov com o espaço de estados $\{0, 1, 2, 3\}$ e a matriz de transição

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

determinaremos se a cadeia de Markov com esta matriz de transição é irredutível.

Solução. Temos

Na Figura 2 note que o estado 0 (ou 1) é acessível do estado 2, mas o estado 2 não é

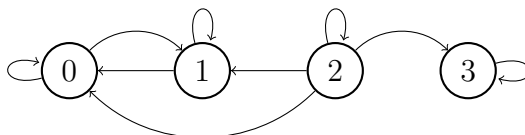


Figura 2: Diagrama de transições para a cadeia de Markov do Exemplo 2

acessível do estado 1. Então os estados não são comunicáveis e a cadeia não é irredutível. \square

Definição 3.5.7. Para cada estado i de uma cadeia de Markov $(X_n)_{n \in \mathbb{N}}$ definimos a probabilidade f_i de que começando no estado i a cadeia vai voltar ao estado i :

$$f_i = P\{\text{existe um instante } n > 0 \text{ tal que } X_n = i \mid X_0 = i\} \quad (3.14)$$

Definição 3.5.8. Estado i é recorrente se $f_i = 1$. Estado i é transiente se $f_i < 1$.

Proposição 3.5.9. Um estado i é recorrente se e somente se, começando do estado i , o número médio de instantes de tempo quando a cadeia está no estado i é infinito.

Demonstração. Se o estado i é recorrente, então a cadeia de Markov, saindo do estado i , necessariamente volta para este estado. Então o número de vezes quando a cadeia está em estado i é infinito. Logo a média deste número é infinita.

Se i é transiente, então existe uma probabilidade positiva de que saindo deste estado i , a cadeia nunca mais vai voltar neste estado (“sucesso”). A probabilidade é igual a $1 - f_i$. Então o número de instantes quando a cadeia está neste estado (incluindo o instante inicial) tem a distribuição geométrica:

$$P\{X_n \text{ exatamente } k \text{ vezes está em estado } i \mid X_0 = i\} = f_i^{k-1}(1 - f_i), \quad k = 1, 2, \dots$$

A média de número de instantes quando a cadeia está no estado i é finita e igual à $1/(1 - f_i)$. \square

Observação 7. Utilizando raciocínio análogo ao da demonstração obtemos que se uma cadeia de Markov tem espaço de estados finito, então pelo menos um dos estados tem que ser recorrente.

Proposição 3.5.10. *Estado i é*

$$\text{recorrente se } \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty,$$

$$\text{transiente se } \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty.$$

Na verdade, ser recorrente ou transiente é propriedade de classe de estados.

Proposição 3.5.11. *Se um estado i é recorrente e $i \leftrightarrow j$, então o estado j é recorrente também.*

Demonstração. Pela Definição 3.5.4, $i \leftrightarrow j$ significa que existem números inteiros k e m tais que $p_{ij}^{(k)} > 0$ e $p_{ji}^{(m)} > 0$. Pelas equações de Chapman-Kolmogorov 3.13, para cada $n \in \mathbb{N}$, temos

$$p_{jj}^{(m+n+k)} \geq p_{ji}^{(m)} p_{ii}^{(n)} p_{ij}^{(k)}. \quad (3.15)$$

Informalmente, desigualdade (3.15) é válida pois a parte esquerda é probabilidade de ir de j para j em $m + n + k$ passos e a parte direita é probabilidade de ir de j para j em $m + n + k$ passos, utilizando caminho que vai de j para i em m passos, depois de i para i em n passos adicionais e depois de i para j em k passos adicionais.

Somando (3.15) em n obtemos

$$\sum_{n=1}^{\infty} p_{jj}^{(m+n+k)} \geq \sum_{n=1}^{\infty} p_{ji}^{(m)} p_{ii}^{(n)} p_{ij}^{(k)} = p_{ji}^{(m)} p_{ij}^{(k)} \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty.$$

Na última equação utilizamos a Proposição 3.5.10 e a recorrência do estado i . \square

Observação 8. Pela Observação 7 e pela Proposição 3.5.11, obtemos que se uma cadeia de Markov irredutível tem espaço de estados finito, então todos os estados são recorrentes.

Exemplo 3. Seja $(X_n)_{n \in \mathbb{N}}$ uma cadeia de Markov com espaço de estados $\{0, 1, 2, 3\}$ e matriz de transição

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Determine quais estados são transientes e quais recorrentes.

Solução. Temos

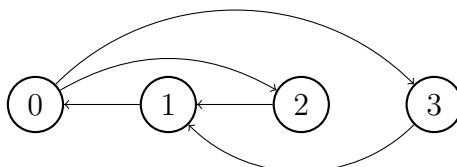


Figura 3: Diagrama de transições para a cadeia de Markov do Exemplo 3

Na Figura 3 é fácil verificar que a cadeia é irredutível. Então, pela Observação 8, todos os estados são recorrentes. \square

Exemplo 4. Seja $(X_n)_{n \in \mathbb{N}}$ uma cadeia de Markov com espaço dos estados $\{0, 1, 2, 3, 4\}$ e matriz de transição

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/2 \end{pmatrix}$$

Determine quais estados são transientes e quais recorrentes.

Solução. Temos

Na Figura 4 a cadeia tem três classes $C_1 = \{0, 1\}$, $C_2 = \{2, 3\}$ e $C_3 = \{4\}$. As primeiras

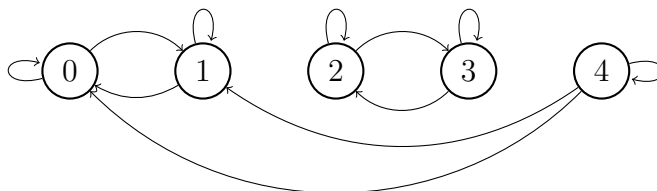


Figura 4: Diagrama de transições para a cadeia de Markov do Exemplo 4

duas são recorrentes e a terceira transiente. \square

3.5.1.3 Ergodicidade e recorrência positiva

Definição 3.5.12. Um estado i tem período d se

$$\text{mdc}\{n \geq 1 : p_{ii}^{(n)} > 0\} = d.$$

Estado com $d = 1$ é aperiódico.

Observação 9. É possível provar que periodicidade é propriedade de classe de estados. Isto é, se i tem período d e $i \leftrightarrow j$, então j também tem período d .

Definição 3.5.13. Se um estado i é recorrente, então ele é recorrente positivo se, começando em i , o tempo médio de retorno a i é finito. Um estado i é ergódico se este estado é aperiódico e recorrente positivo.

Observação 10. É possível provar que recorrência positiva é propriedade de classe e que numa cadeia de Markov com espaço de estados finito todos estados recorrentes são recorrentes positivos.

Definição 3.5.14. Uma cadeia de Markov é ergódica se todos os estados dela são ergódicos. (Existem definições análogas para propriedade de estado ser transiente, recorrente etc.)

3.5.2 Caso contínuo

Definição 3.5.15. Um processo estocástico $(X_t)_{t \in [0, +\infty)}$, tendo como espaço de estados um conjunto enumerável E , é dito uma cadeia de Markov em tempo contínuo se para todos $s, t \in [0, +\infty)$ e para todos estados $x, y, z(u) \in E, 0 \leq u < s$, temos

$$P\{X(t+s) = y | X(s) = x, X(u) = z(u), 0 \leq u < s\} = P\{X(t+s) = y | X(s) = x\}. \quad (3.16)$$

Definição 3.5.16. Um processo estocástico $(X_t)_{t \in [0, +\infty)}$, tendo como espaço de estados um conjunto enumerável E , é dito uma cadeia de Markov em tempo contínuo se cada instante do tempo que ele entra em qualquer estado $x \in E$

- i. a quantidade de tempo que ele passa neste estado antes de fazer transição para outro estado tem distribuição exponencial com algum parâmetro que vamos denotar por v_x .
- ii. quando o processo sai do estado x ele entra no estado y com alguma probabilidade

que vamos denotar por p_{xy} que claramente satisfaz

$$p_{xx} = 0 \text{ e } \sum_{y \in E} p_{xy} = 1.$$

De acordo com a Seção 6.4 de [3] temos,

$$P_{xy}(t) = P\{X(t+s) = y | X(s) = x\} \quad (3.17)$$

indica que a probabilidade de um processo que está no estado x vai passar para o estado y com tempo t . Essa quantidade é comumente chamada de probabilidade de transição de uma cadeia de markov em tempo contínuo.

Observação 11. Para qualquer par de estados x e y , seja

$$q_{xy} = v_x p_{xy}$$

Desde que v_x seja uma taxa que marca a transição do estado x e p_{xy} a probabilidade que essa transição é feita para o estado y , isso segue que q_{xy} é a taxa de transição do estado x para o estado y . As quantidades q_{xy} são chamadas de taxa de transição instantânea. Desde que,

$$v_x = \sum_y v_x p_{xy} = \sum_y q_{xy} \quad (3.18)$$

e

$$p_{xy} = \frac{q_{xy}}{v_x} = \frac{q_{xy}}{\sum_y q_{xy}} \quad (3.19)$$

O Lema 6.2 de [3] nos diz que

a)

$$\lim_{h \rightarrow 0^+} \frac{1 - P_{xx}(h)}{h} = v_x$$

b)

$$\lim_{h \rightarrow 0^+} \frac{1 - P_{xy}(h)}{h} = q_{xy} \quad \text{quando } x \neq y$$

3.6 Processo de Nascimento e Morte

Os Processos de Nascimento e Morte são processos estocásticos muito utilizados na descrição de Sistemas de Filas, dada a sua particular estrutura: as transições de um estado qualquer só são possíveis para estados “vizinhos” (i.e., de um estado i para os estados $i + 1$ ou $i - 1$ se $i \geq 1$).

Adotando-se um espaço de estados $X = 0, 1, \dots$ e considerando que cada estado representa um certo “nível de população” (exemplo: número de clientes numa loja, número de mensagens num coletor de chamadas, número de produtos a processar, etc.), tais transições podem ser facilmente interpretadas.

A transição do estado i para o estado $i + 1$ será um “nascimento” (por exemplo, chegada de um cliente), uma vez que significa um aumento do “nível da população”. Enquanto que a transição do estado i para o estado $i - 1$ será uma “morte” (por exemplo, partida de um cliente), por significar um decréscimo do “nível da população” e não é possível quando $X = 0$, ou seja, nenhum cliente no sistema.

Definição 3.6.1. *O processo de nascimento e morte X_t é uma Cadeia de Markov em tempo contínuo com espaço de estados $N = \{0, 1, 2, \dots\}$ e taxa de transição instantânea*

$$q_{xy} = \begin{cases} \lambda_x & \text{se } y = x + 1 \text{ e } x \geq 0, \\ \mu_x & \text{se } y = x - 1 \text{ e } x \geq 1. \end{cases} \quad (3.20)$$

A taxa de transição instantânea igual a λ_x representa a chegada de um cliente ao sistema, pois é a taxa de transição de x para y se $y = x + 1$. Já a taxa de transição instantânea igual a μ_x representa a saída de um cliente do sistema, pois é a taxa de transição de x para y se $y = x - 1$.

Na seção 6.5 de [3] é encontrado o seguinte teorema:

Teorema 3.6.2. *Seja $\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} < +\infty$. Então a distribuição estacionária para o processo de nascimento e morte é*

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}} \quad (3.21)$$

e

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} p_0, \quad n \geq 1. \quad (3.22)$$

Note que nesta monografia não colocamos a definição formal de distribuição estacionária e as propriedades dela. O fato de 3.22 ser uma distribuição estacionária, informalmente, significa que depois de algum tempo inicial, em cada momento de tempo a probabilidade de encontrar n pessoas no sistema é aproximadamente p_n , e p_n também é aproximadamente a proporção de tempo em que tinham n clientes no sistema.

3.6.1 Exemplo de um nó simples

Nesta seção será abordado o modelo $M/M/1$, ou seja, sistema com apenas um nó simples onde tanto as chegadas quanto o atendimento são markovianas. Os momentos de chegadas dos clientes ao sistema formam um processo de Poisson Homogêneo com taxa λ e o tempo de serviço é distribuído exponencialmente com taxa μ , como é ilustrado na Figura 5.

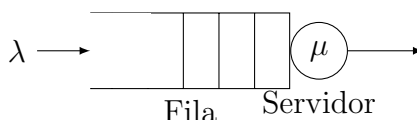


Figura 5: Representação do sistema com um nó

Nesse modelo, $\lambda_n = \lambda$, $n = 0, 1, \dots$, e $\mu_n = \mu$, $n = 1, 2, \dots$, logo a partir de (3.22) e (3.21), temos

$$\begin{aligned} p_0 &= \frac{1}{1 + \sum_{n=1}^{\infty} (\lambda/\mu)^n} \\ &= 1 - \lambda/\mu \end{aligned} \quad (3.23)$$

e

$$p_n = \left(\frac{\lambda}{\mu}\right)^n p_0, \quad n \geq 1, \quad (3.24)$$

desde que $\frac{\lambda}{\mu} < 1$. A partir de p_n , temos que o tamanho de fila tem distribuição Geométrica com parâmetro $1 - \rho$, onde $\rho = \frac{\lambda}{\mu}$.

A seguir, vamos calcular as medidas de performance para esse modelo.

Utilização

A utilização é a média de uma variável que toma dois valores: 0 e 1, com probabilidades p_0 e $1 - p_0$, então

$$\begin{aligned} U &= 1 - p_0 \\ &= 1 - \left(1 - \frac{\lambda}{\mu}\right) \end{aligned}$$

Logo,

$$U = \frac{\lambda}{\mu}. \quad (3.25)$$

Tamanho da fila

Usando (3.23) e (3.24), temos que o tamanho médio da fila Q é dado pela média de uma Geométrica com parâmetro $1 - \rho$, então

$$Q = \frac{\rho}{1 - \rho}.$$

Logo,

$$Q = \frac{\lambda}{\mu - \lambda}. \quad (3.26)$$

Tamanho da fila de espera

Seja p'_k a probabilidade de ter k pessoas na fila de espera, então

$$\begin{aligned} p'_0 &= p_0 + p_1 \\ p'_1 &= p_2 \\ p'_2 &= p_3 \\ &\vdots \\ p'_n &= p_{n+1} \\ &\vdots \end{aligned}$$

Temos que o tamanho médio da fila de espera L é

$$\begin{aligned} L &= \sum_{i=1}^{\infty} p'_i i \\ &= \sum_{i=1}^{\infty} p_{i+1} i \\ &= \sum_{i=2}^{\infty} p_i (i - 1) \\ &= \sum_{i=2}^{\infty} p_i i - \sum_{i=2}^{\infty} p_i \\ &= \left(\frac{\rho}{1 - \rho} - p_1 \right) - (1 - p_1 - p_0) \\ &= \frac{\rho}{1 - \rho} - (1 - p_0) \\ &= \frac{\rho}{1 - \rho} - (1 - (1 - \rho)) \\ &= \frac{\rho}{1 - \rho} - \rho. \end{aligned}$$

Logo,

$$L = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu}. \quad (3.27)$$

Tempo de permanência no sistema

No Capítulo 4 de [1], a primeira lei de Little afirma que o tamanho médio da fila Q é igual à taxa média de chegada multiplicada pelo tempo médio gasto no sistema R ,

$$Q = \lambda R.$$

Então podemos encontrar o valor do tempo de permanência:

$$\begin{aligned} R &= \frac{Q}{\lambda} \\ R &= \frac{\frac{\lambda}{\mu - \lambda}}{\lambda}. \end{aligned}$$

Logo,

$$R = \frac{1}{\mu - \lambda}. \quad (3.28)$$

Tempo de espera

O tempo de serviço é o tempo gasto por cliente no atendimento. Mais formalmente, o tempo médio de serviço S pode ser obtido a partir do inverso da taxa de serviço μ . Logo,

$$S = \frac{1}{\mu}.$$

Seja W o tempo médio gasto na fila de espera. O tempo de permanência no sistema é a soma do tempo de espera e do tempo de serviço, então:

$$\begin{aligned} R &= W + S \\ W &= \frac{1}{\mu - \lambda} - S. \end{aligned}$$

Logo,

$$W = \frac{\lambda}{\mu(\mu - \lambda)}. \quad (3.29)$$

3.6.2 Exemplo de um nó com servidor duplo

Se adicionarmos um servidor idêntico à uma mesma fila de um sistema com um nó, teremos um sistema de filas de um nó com servidor duplo, que será o exemplo abordado nessa seção. Sistemas de múltiplos servidores são tipicamente encontrados em caixas de banco, correio, call centers e CPUs com múltiplos processadores.

Os momentos de chegadas dos clientes ao sistema formam um processo de Poisson

Homogêneo com taxa λ e o tempo de serviço é distribuído exponencialmente com taxa μ para cada servidor, como é ilustrado na Figura 6.

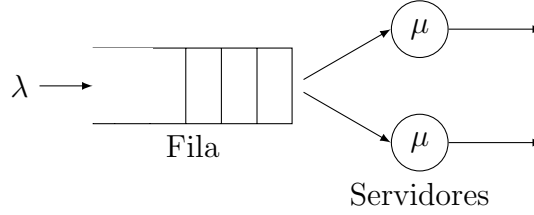


Figura 6: Sistema de um nó com servidor duplo

Nesse modelo, $\lambda_n = \lambda$, se $n \geq 1$, $\mu_1 = \mu$ e $\mu_n = 2\mu$, se $n \geq 2$. Assim, podemos desenvolver (3.21) da seguinte maneira

$$\begin{aligned}
 p_0 &= \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n}} = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda^n}{\mu(2\mu)^{n-1}}} \\
 &= \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda}{\mu} \left(\frac{\lambda}{2\mu}\right)^{n-1}} = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda}{\mu} \left(\frac{\lambda}{2\mu}\right)^n \left(\frac{\lambda}{2\mu}\right)^{-1}} \\
 &= \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda}{\mu} \frac{2\mu}{\lambda} \left(\frac{\lambda}{2\mu}\right)^n} = \frac{1}{1 + 2 \sum_{n=1}^{\infty} \left(\frac{\lambda}{2\mu}\right)^n} \quad (3.30) \\
 &= \frac{1}{1 + 2 \left(\frac{\frac{\lambda}{2\mu}}{1 - \frac{\lambda}{2\mu}}\right)} = \frac{1}{1 + 2 \left(\frac{\lambda}{2\mu - \lambda}\right)} \\
 &= \frac{2\mu - \lambda}{2\mu - \lambda + 2\lambda}.
 \end{aligned}$$

Logo,

$$p_0 = \frac{2\mu - \lambda}{2\mu + \lambda}. \quad (3.31)$$

e a partir de (3.22)

$$p_n = \frac{\lambda}{\mu} \left(\frac{\lambda}{2\mu}\right)^{n-1} p_0, \quad n \geq 1, \quad (3.32)$$

desde que $\frac{\lambda}{2\mu} < 1$.

A seguir, vamos calcular as medidas de performance para esse modelo.

Utilização

A utilização é a média de uma variável que toma três valores: 0, 1 e 2, com probabilidades

p_0 , p_1 e $1 - p_0 - p_1$, respectivamente, dividida pelo número de servidores, então

$$\begin{aligned}
 U &= \frac{p_1 + 2(1 - p_0 - p_1)}{2} \\
 &= \frac{2 - 2p_0 - p_1}{2} \\
 &= 1 - p_0 - \frac{p_1}{2} \\
 &= 1 - p_0 - \frac{\lambda p_0}{2\mu} \\
 &= 1 - p_0 \left(1 + \frac{\lambda}{2\mu}\right) \\
 &= 1 - \frac{2\mu - \lambda}{2\mu + \lambda} \left(1 + \frac{\lambda}{2\mu}\right) \\
 &= 1 - \frac{\left(1 - \frac{\lambda}{2\mu}\right)}{\left(1 + \frac{\lambda}{2\mu}\right)} \left(1 + \frac{\lambda}{2\mu}\right) \\
 &= 1 - \left(1 - \frac{\lambda}{2\mu}\right)
 \end{aligned}$$

Logo,

$$U = \frac{\lambda}{2\mu} \quad (3.33)$$

Tamanho da fila

Através da média de uma variável com distribuição geométrica e parâmetro p , sabemos que

$$\sum_{n=1}^{\infty} n(1-p)^n(p) = \frac{1-p}{p},$$

substituindo p por $1-x$, temos

$$\sum_{n=1}^{\infty} nx^n(1-x) = \frac{x}{1-x},$$

agora, vamos dividir a igualdade por x

$$\sum_{n=1}^{\infty} nx^{n-1}(1-x) = \frac{1}{1-x},$$

substituindo x por $\frac{\lambda}{2\mu}$, temos

$$\sum_{n=1}^{\infty} n \left(\frac{\lambda}{2\mu}\right)^{n-1} \left(1 - \frac{\lambda}{2\mu}\right) = \frac{1}{1 - \frac{\lambda}{2\mu}},$$

como queremos encontrar a média de p_n (tamanho médio da fila), vamos multiplicar

ambos os lados da igualdade por $p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right)$

$$p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \sum_{n=1}^{\infty} n \left(\frac{\lambda}{2\mu} \right)^{n-1} \left(1 - \frac{\lambda}{2\mu} \right) = p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right),$$

então

$$\sum_{n=1}^{\infty} n \frac{\lambda}{\mu} \left(\frac{\lambda}{2\mu} \right)^{n-1} p_0 = p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right),$$

$$\sum_{n=1}^{\infty} np_n = p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right),$$

ou seja

$$\begin{aligned} Q &= p_0 \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right) \\ &= \left(\frac{2\mu - \lambda}{2\mu + \lambda} \right) \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right) \\ &= \left(\frac{\frac{1}{1 + \frac{\lambda}{2\mu}}}{\frac{1}{1 - \frac{\lambda}{2\mu}}} \right) \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right) \left(\frac{1}{1 - \frac{\lambda}{2\mu}} \right) \\ &= \left(\frac{1}{1 + \frac{\lambda}{2\mu}} \right) \left(\frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{2\mu}} \right). \end{aligned}$$

Logo,

$$Q = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu} \right)^2} \quad (3.34)$$

Tamanho da fila de espera

Seja p'_k a probabilidade de ter k pessoas na fila de espera, então

$$p'_0 = p_0 + p_1 + p_2$$

$$p'_1 = p_3$$

$$p'_2 = p_4$$

⋮

$$p'_n = p_{n+2}$$

⋮

.

Assim, temos que o tamanho médio da fila de espera L é

$$\begin{aligned}
L &= \sum_{i=1}^{\infty} p'_i i \\
&= \sum_{i=1}^{\infty} p_{i+2} i \\
&= \sum_{i=3}^{\infty} p_i (i-2) \\
&= \sum_{i=3}^{\infty} (p_i i - 2p_i) \\
&= \sum_{i=3}^{\infty} p_i i - 2 \sum_{i=3}^{\infty} p_i \\
&= (Q - 2p_2 - p_1) - 2(1 - p_2 - p_1 - p_0) \\
&= Q - 2 + 2p_0 + p_1 \\
&= Q - 2 + 2p_0 + \frac{\lambda}{\mu} p_0 \\
&= Q - 2 \left(1 - p_0 - \frac{\lambda}{2\mu} p_0 \right) \\
&= Q - 2 \left[1 - p_0 \left(1 + \frac{\lambda}{2\mu} \right) \right] \\
&= Q - 2 \left[1 - \left(\frac{2\mu - \lambda}{2\mu + \lambda} \right) \left(1 + \frac{\lambda}{2\mu} \right) \right] \\
&= Q - 2 \left[1 - \left(\frac{1 - \frac{\lambda}{2\mu}}{1 + \frac{\lambda}{2\mu}} \right) \left(1 + \frac{\lambda}{2\mu} \right) \right] \\
&= Q - 2 \left[1 - \left(1 - \frac{\lambda}{2\mu} \right) \right] \\
&= Q - \frac{\lambda}{\mu}
\end{aligned}$$

Logo,

$$L = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2} - \frac{\lambda}{\mu}. \quad (3.35)$$

Tempo de permanência no sistema

Assim como no exemplo anterior, de um nó simples, o tempo médio de permanência gasto no sistema será calculado através da primeira lei de Little, ou seja, através de $Q = \lambda R$.

Logo

$$R = \frac{Q}{\lambda}.$$

Logo,

$$R = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2}. \quad (3.36)$$

Tempo de espera

Novamente como no exemplo anterior, o tempo médio gasto na fila de espera W será obtido através do tempo de permanência no sistema menos o tempo de serviço. Ou seja,

$$W = R - S$$

Como $S = \frac{1}{\mu}$, temos

$$W = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2} - \frac{1}{\mu}. \quad (3.37)$$

3.7 Redes de Jackson

Os exemplos mais simples de sistemas de filas impõem, juntamente com algumas generalizações, a restrição de que os clientes só podem se mover ao longo de um caminho, ou seja, não pode haver retornos (*feed-backward*).

Consideramos agora um sistema de filas de modo mais geral, onde existe um número finito K de filas individuais, os nós do sistema, em que os clientes chegam de fontes externas de acordo com processos de Poisson independentes com intensidades $\alpha_1, \dots, \alpha_K$. Um cliente após ser servido no nó k vai para o nó l com probabilidade γ_{kl} e sai do sistema com probabilidade $\gamma_{k0} = 1 - \sum_{l=1}^K \gamma_{kl}$.

Um sistema de filas é formado por nós, onde cada nó contém um ou mais servidores e uma fila. Nesta seção serão considerados apenas os casos de um único servidor exponencial em cada nó com taxas de serviços dadas por $\delta_1, \dots, \delta_K$.

Existem dois tipos principais de sistemas de filas, os abertos, onde clientes entram e saem do sistema, e fechados, onde os clientes só podem se mover internamente no sistema. Assim o número total de clientes no sistema fechado não varia com o tempo. Em uma rede aberta é necessário que o sistema permaneça estável, ou seja, a quantidade de clientes que chegam ao sistema não deve ser muito maior do que a quantidade de saídas em um determinado período.

Temos que $X_t^{(k)}$ denota o número de clientes no nó k no tempo t e que $X_t = (X_t^{(1)}, \dots, X_t^{(k)})$ é definida como uma cadeia de Markov em tempo contínuo com espaço

de estados \mathbb{N}^K e taxa de transição instantânea

$$q_{xy} = \begin{cases} \delta_k \gamma_{kl} & \text{se } y_k = x_k - 1, y_l = x_l + 1, k = 1, \dots, K, l = 1, \dots, K \text{ e } k \neq l, \\ \alpha_k & \text{se } y_k = x_k + 1, k = 1, \dots, K, \\ \delta_k \gamma_{k0} & \text{se } y_k = x_k - 1, k = 1, \dots, K. \end{cases} \quad (3.38)$$

A taxa de transição instantânea igual a $\delta_k \gamma_{kl}$ representa a passagem de um cliente após ser atendido nó k para o nó l , pois é a taxa de transição de x para y se $y_k = x_k - 1$, $y_l = x_l + 1$ e $k \neq l$. Já a taxa de transição instantânea igual a α_k representa a chegada de um cliente ao nó k , pois é a taxa de transição de x para y se $y_k = x_k + 1$. E um cliente deixando o sistema após ser atendido no nó k é representado pela taxa de transição instantânea $\delta_k \gamma_{k0}$, taxa de transição de $y_k = x_k - 1$. Assim, apenas o primeiro tipo de transição ocorre num sistema fechado.

Precisamos fazer uma definição adequada da taxa de transferência β_k no nó k , isto é, a taxa comum dos processos de entrada e de saída (esta não precisa ser Poisson, mas a taxa deve existir em termos de médias). A taxa de entrada é a soma das α_k taxas de chegadas externas e as taxas de chegadas internas dos nós $l \neq k$. Mas os clientes deixam o nó l com taxa β_l e, em seguida, vão para k com probabilidade γ_{lk} , de modo que temos as equações de trânsito:

$$\beta_k = \alpha_k + \sum_{l=1}^K \beta_l \gamma_{lk}, \quad k = 1, \dots, K. \quad (3.39)$$

Temos noções de irredutibilidade (Definição 3.5.6) e ergodicidade (Definição 3.5.14) definidos para cadeia de Markov em tempo discreto, e nesta seção vamos usar elas para cadeia de Markov em tempo contínuo. Cadeia de Markov em tempo contínuo é irredutível se a cadeia de Markov em tempo discreto correspondente a ela (i.e. definida no mesmo espaço de estados e que tem como probabilidades de transição aqueles p_{ij} da Definição 3.5.15) é irredutível.

Vamos supor que cada nó k pode ter tanto entrada externa quanto saída externa (possivelmente através de outros nós), ou seja, para cada k :

- i. $\alpha_k > 0$ ou algum $\alpha_{l_1} \gamma_{l_1 l_2} \dots \gamma_{l_r k} > 0$, e
- ii. ou $\gamma_{k0} > 0$ ou algum $\gamma_{k k_1} \gamma_{k_1 k_2} \dots \gamma_{k_n 0} > 0$.

Isto implica a irredutibilidade de $\{X_t\}_{t \geq 0}$.

Proposição 3.7.1. *As equações de trânsito (3.39) têm uma única solução não negativa $(\beta_1, \dots, \beta_k)$. Ela satisfaz $0 < \beta_k < \infty$.*

Proposição 3.7.2. *Estamos supondo que a taxa de utilização $\rho_k = \beta_k/\delta_k < 1$ para todo $k = 1, \dots, K$. Assim, $\{X_t\}$ é ergódico com distribuição estacionária π dada por*

$$\pi_n = \pi_{n_1 \dots n_K} = \prod_{k=1}^K (1 - \rho_k) \rho_k^{n_k}, \quad (3.40)$$

ou seja, cada tamanho de fila tem distribuição Geométrica com parâmetro $1 - \rho_k$ e os tamanhos de filas são independentes. Veja o Teorema 2.2 de [4].

3.7.1 Exemplo de um nó simples

O modelo abordado nessa seção será o mesmo da 3.6.1, que está ilustrado na Figura 5, porém será desenvolvido através de Redes de Jackson. Temos assim, um exemplo de Rede de Jackson com $K = 1$, $\alpha_1 = \lambda$ e $\gamma_{10} = \mu$.

A equação de trânsito (3.39) é dada por $\beta_1 = \lambda$, e a solução dessa equação é $\beta_1 = \lambda$.

3.7.2 Medidas de performance para Redes de Jackson

Para $k = 1, \dots, K$, supondo que a Rede de Jackson satisfaz as condições do Teorema 3.7.2 e seguindo o Capítulo 5 do [1], podemos calcular as medidas de performance para cada nó k , para cada uma das equações (3.25)-(3.29) colocando

$$\lambda = \beta_k/V_k \text{ e } \mu = \delta_k/V_k, \text{ onde } V_k = \frac{\beta_k}{\sum_{i=1}^K \alpha_i}. \quad (3.41)$$

Logo,

$$U = \frac{\lambda}{\mu} = \frac{\beta_k/V_k}{\delta_k/V_k} = \frac{\beta_k}{\delta_k}, \quad (3.42)$$

$$Q = \frac{\lambda}{\mu - \lambda} = \frac{\beta_k/V_k}{\delta_k/V_k - \beta_k/V_k} = \frac{\beta_k}{\delta_k - \beta_k}, \quad (3.43)$$

$$L = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\beta_k/V_k}{\delta_k/V_k - \beta_k/V_k} - \frac{\beta_k/V_k}{\delta_k/V_k} = \frac{\beta_k}{\delta_k - \beta_k} - \frac{\beta_k}{\delta_k}, \quad (3.44)$$

$$R = \frac{1}{\mu - \lambda} = \frac{1}{\delta_k/V_k - \beta_k/V_k} = \frac{V_k}{\delta_k - \beta_k} = \frac{\beta_k}{(\delta_k - \beta_k) \sum_{i=1}^K \alpha_i}, \quad (3.45)$$

$$W = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\beta_k/V_k}{\delta_k/V_k(\delta_k/V_k - \beta_k/V_k)} = \frac{\beta_k V_k}{\delta_k(\delta_k - \beta_k)} = \frac{\beta_k^2}{\delta_k(\delta_k - \beta_k) \sum_{i=1}^K \alpha_i}. \quad (3.46)$$

4 Modelos Utilizados

Para os modelos que serão analisados neste trabalho, serão consideradas algumas características específicas:

- A população será considerada infinita.
- O sistema tem capacidade infinita, ou seja, não há rejeição de clientes que chegam ao sistema nem desistências.
- Cada cliente chega individualmente ao sistema, ou seja, não consideramos chegadas em grupo.
- Não consideramos nenhum tipo de desistência por parte dos clientes.
- Consideramos que a fila é do tipo FIFO, conseqüentemente, os clientes que chegam primeiro ao sistema são atendidos primeiro e não consideramos nenhum tipo de prioridade.
- Não será permitido que um servidor fique ocioso (livre) quando há trabalho a ser realizado, quando houver fila. Assim sendo, o primeiro cliente na fila será atendido assim que algum dos servidores terminar de processar o trabalho que estiver realizando.

Outras características, como a taxa de chegada, a taxa de atendimento, o número de servidores e a configuração dos servidores, serão apresentadas especificamente em cada modelo descritos a seguir.

4.1 Modelo 1 - Dois servidores em paralelo e fila única

O Modelo 1 que será abordado neste trabalho corresponde ao apresentado na Figura 7. Há apenas uma classe de clientes, os momentos de chegadas dos clientes ao sistema

formam um processo de Poisson Homogêneo com taxa λ e o tempo de serviço de cada servidor é distribuído exponencialmente com taxa μ . Para a análise dos resultados da simulação e do PDQ, serão utilizados $\lambda = 0,75$ e $\mu = 1$.

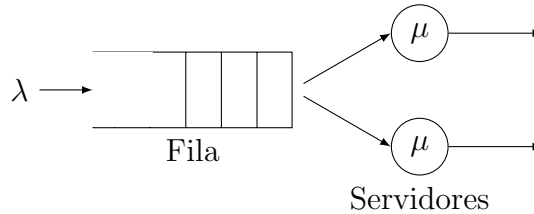


Figura 7: Sistema de fila única e dois servidores homogêneos

4.1.1 Cadeias de Markov - Modelo 1

Usando a Definição 3.5.16 junto com a Observação 11, definimos, X_t como uma cadeia de Markov em tempo contínuo com espaço de estados $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, e taxa de transição instantânea

$$q_{xy} = \begin{cases} \lambda & \text{se } y = x + 1, \\ 2\mu & \text{se } x \geq 2 \text{ e } y = x - 1, \\ \mu & \text{se } x = 1 \text{ e } y = 0. \end{cases} \quad (4.1)$$

Essa definição é ilustrada na Figura 8.

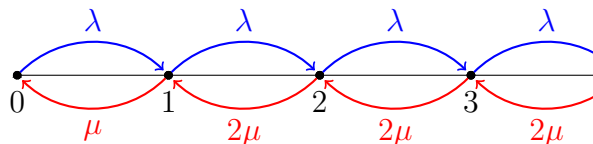


Figura 8: Diagrama de transição de estados para o Modelo 1.

Nesse modelo o estado inicial é 0, que é o número de clientes dentro do sistema quando ele é iniciado. A única alteração de estado que pode ocorrer nesse momento é para o estado 1 com probabilidade $\frac{\lambda}{\lambda}$.

Quando o sistema tem apenas um cliente, é representado pelo estado 1, que tem como possíveis transições os estados 0 e 2. Ou seja, para haver alguma transição de estado é necessário que chegue alguém ao sistema (fazendo assim a transição para o estado 2) ou que esse cliente termine de ser atendido (voltando ao estado inicial). A probabilidade de

passar do estado 1 para o 2 é $\frac{\lambda}{\mu+\lambda}$, e para voltar ao estado 0 (inicial) a probabilidade é $\frac{\mu}{\mu+\lambda}$.

O estado 2, quando o sistema tem dois clientes, tem como possíveis transições os estados 1 e 3. Ou seja, para haver alguma transição de estado é necessário que chegue alguém ao sistema (fazendo assim a transição para o estado 3) ou que algum dos clientes em atendimento termine de ser atendido e saia do sistema (fazendo a transição para o estado 1). A probabilidade de passar do estado 2 para o 3 é $\frac{\lambda}{2\mu+\lambda}$, e para passar para o estado 1 a probabilidade é $\frac{2\mu}{2\mu+\lambda}$.

A partir do estado 2, todos os outros estados seguem as mesmas possíveis transições: chegar algum cliente ao sistema ou um dos dois que estão em atendimento sair do sistema. E seguem também as mesmas probabilidades de transições.

Como a capacidade do sistema é ilimitada e o tamanho da população é infinito, não existe um estado final para o diagrama de transição desse modelo. Isso é equivalente a afirmar que independente do estado sempre haverá possibilidade de transição.

A taxa de transição instantânea igual a λ representa a chegada de um cliente ao sistema, pois é a taxa de transição de x para y se $y = x + 1$. Já a taxa de transição instantânea igual a μ representa a saída do único cliente no sistema, pois é a taxa de transição de x para y se $x = 1$ e $y = 0$, ou seja, μ é a taxa de transição de 1 para 0. E a saída de um cliente do sistema, quando este tem dois ou mais clientes, é representada pela taxa de transição instantânea 2μ , taxa de transição de $x \geq 2$ e $y = x - 1$.

Essas taxas de transição instantâneas já haviam sido observadas nas setas do diagrama de transição de estados. E as probabilidades dessas transições, que também já foram apresentadas, são calculadas seguindo a Observação 11, ou seja, a probabilidade de transição de um estado x para um estado y (p_{xy}) é a taxa de transição desses estados (q_{xy}) dividido pela soma das taxas de todas as possíveis transições do estado x para outros estados (v_x).

4.1.2 Medidas de performance - Modelo 1

Usando (3.33)-(3.37), vamos calcular as medidas de performance para esse modelo considerando $\lambda = 0,75$ e $\mu = 1$, que serão os valores utilizados na simulação e no PDQ, temos:

$$U = \frac{\lambda}{2\mu} = \frac{0,75}{2} = 0,375$$

$$Q = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2} = \frac{0,75}{1 - \left(\frac{0,75}{2}\right)^2} = 0,8727$$

$$L = \frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2} - \frac{\lambda}{\mu} = \frac{0,75}{1 - \left(\frac{0,75}{2}\right)^2} - 0,75 = 0,1227$$

$$R = \frac{\frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2}}{\lambda} = \frac{\frac{0,75}{1 - \left(\frac{0,75}{2}\right)^2}}{0,75} = 1,1636$$

$$W = \frac{\frac{\frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{2\mu}\right)^2}}{\lambda} - \frac{1}{\mu} = \frac{\frac{0,75}{1 - \left(\frac{0,75}{2}\right)^2}}{0,75} - 1 = 0,1636$$

As informações do Modelo 1 são passadas para o PDQ da seguinte maneira:

```
lambda = 0.75
mu = 1
library("pdq")
Init("OpenCenter")
CreateOpen("work", lambda)
CreateMultiNode(2, "server", CEN, MSQ)
SetDemand("server", "work", 1/mu)
Solve(CANON)
Report()
```

Os resultados obtidos com o PDQ são os mesmo que foram obtidos analiticamente através do processo de Nascimento e Morte.

4.2 Modelo 2 - Dois servidores em série

O Modelo 2 que será abordado neste trabalho corresponde ao apresentado na Figura 9. Há apenas uma classe de clientes, os momentos de chegadas dos clientes ao sistema formam um processo de Poisson Homogêneo com taxa λ . O modelo é formado por dois

servidores heterogêneos em série e o tempo de serviço de cada servidor é distribuído exponencialmente com taxa μ_1 e μ_2 , respectivamente. Para a análise dos resultados da simulação e do PDQ, serão utilizados $\lambda = 0,75$, $\mu_1 = 0,25^{-1}$ e $\mu_2 = 0,75^{-1}$.

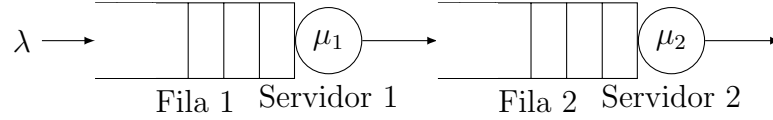


Figura 9: Sistema de dois servidores heterogêneos em série

O Modelo 2 é um exemplo de Jackson Networks com $K = 2$, $\alpha_1 = \lambda$, $\alpha_2 = 0$, $\delta_1 = \mu_1$ e $\delta_2 = \mu_2$. Como nesse modelo não há possibilidades de retorno nem de sair do servidor 1 para fora do sistema, as probabilidades de transições são:

$$\begin{aligned}\gamma_{10} &= 0 \\ \gamma_{12} &= 1 \\ \gamma_{20} &= 1 \\ \gamma_{21} &= 0.\end{aligned}$$

4.2.1 Cadeias de Markov - Modelo 2

Usando a Definição 3.5.16 junto com a Observação 11, definimos, X_t como uma cadeia de Markov em tempo contínuo com espaço de estados \mathbb{N}^2 , e taxa de transição instantânea

$$q_{xy} = \begin{cases} \lambda & \text{se } y_1 = x_1 + 1 \text{ e } y_2 = x_2, \\ \mu_1 & \text{se } x_1 \geq 1 \text{ e } y_1 = x_1 - 1 \text{ e } y_2 = x_2 + 1, \\ \mu_2 & \text{se } x_2 \geq 1 \text{ e } y_2 = x_2 - 1 \text{ e } y_1 = x_1. \end{cases} \quad (4.2)$$

Essa definição é ilustrada na Figura 10.

4.2.2 Equações de trânsito - Modelo 2

As equações de trânsito (3.39) são dadas por

$$\begin{cases} \beta_1 = \lambda, \\ \beta_2 = \beta_1. \end{cases} \quad (4.3)$$

$$W = \frac{\beta_1^2}{\delta_1(\delta_1 - \beta_1)\lambda} = \frac{(0,75)^2}{(4(4 - 0,75))0,75} = 0,06$$

E para o segundo nó, temos:

$$U = \frac{\beta_2}{\delta_2} = \frac{0,75}{1,33} = 0,5625$$

$$Q = \frac{\beta_2}{\delta_2 - \beta_2} = \frac{0,75}{1,33 - 0,75} = 1,29$$

$$L = \frac{\beta_2}{\delta_2 - \beta_2} - \frac{\beta_2}{\delta_2} = \frac{0,75}{1,33 - 0,75} - \frac{0,75}{1,33} = 0,72$$

$$R = \frac{\beta_2}{(\delta_2 - \beta_2)\lambda} = \frac{0,75}{(1,33 - 0,75)0,75} = 1,72$$

$$W = \frac{\beta_2^2}{\delta_2(\delta_2 - \beta_2)\lambda} = \frac{0,75^2}{1,33(1,33 - 0,75)0,75} = 0,97$$

As informações do Modelo 2 são passadas para o PDQ da seguinte maneira:

```
library("pdq")
lambda = 0.75
mu = c(1/0.25, 1/0.75)
Init("Jackson Network in PDQ-R")
wname <- c("Traffic")
rname <- c("Router1", "Router2")
CreateOpen(wname, lambda)
SetWUnit("Msg")
SetTUnit("Sec")
for (i in 1:length(rname)) {
  CreateNode(rname[i], CEN, FCFS)
  SetDemand(rname[i], wname, 1/mu[i])
}
Solve(CANON)
Report()
```

Os resultados obtidos com o PDQ são os mesmo que foram obtidos analiticamente

através da Rede de Jackson.

4.3 Modelo 3 - Rede de Comunicações

O Modelo 3 que será abordado neste trabalho corresponde ao apresentado na Figura 11. Há apenas uma classe de clientes, os momentos de chegadas dos clientes ao sistema formam um processo de Poisson Homogêneo com taxa $\lambda = 0,5$. O modelo é formado por 3 servidores que formam uma rede de comunicações e o tempo de serviço de cada servidor é distribuído exponencialmente com taxa $\mu_1 = 1$, $\mu_2 = 0,5$ e $\mu_3 = 1$, respectivamente.

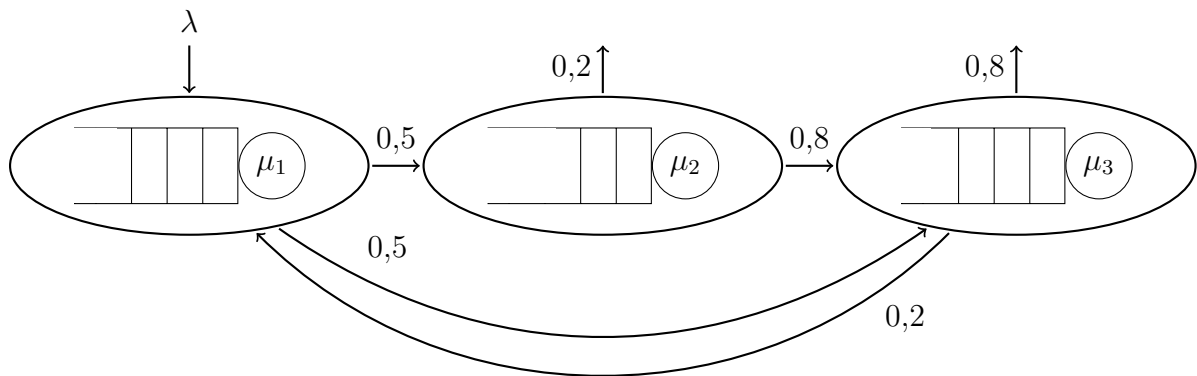


Figura 11: Sistema de uma rede de comunicações

O Modelo 3 é um exemplo de Rede de Jackson com $K = 3$, $\alpha_1 = \lambda$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\delta_1 = 1$, $\delta_2 = 0,5$ e $\delta_3 = 1$. As probabilidades de transições são:

$$\gamma_{10} = 0$$

$$\gamma_{12} = 0,5$$

$$\gamma_{13} = 0,5$$

$$\gamma_{20} = 0,2$$

$$\gamma_{21} = 0$$

$$\gamma_{23} = 0,8$$

$$\gamma_{30} = 0,8$$

$$\gamma_{31} = 0,2$$

$$\gamma_{32} = 0.$$

4.3.1 Cadeias de Markov - Modelo 3

Usando a Definição 3.5.16 junto com a Observação 11, definimos, X_t como uma cadeia de Markov em tempo contínuo com espaço de estados \mathbb{N}^3 , e taxa de transição instantânea

$$q_{xy} = \begin{cases} 0,5\mu_1 & \text{se } y_1 = x_1 - 1, y_2 = x_2 + 1 \text{ e } y_3 = x_3, \\ 0,5\mu_1 & \text{se } y_1 = x_1 - 1, y_2 = x_2 \text{ e } y_3 = x_3 + 1, \\ 0,8\mu_2 & \text{se } y_1 = x_1, y_2 = x_2 - 1 \text{ e } y_3 = x_3 + 1, \\ 0,2\mu_3 & \text{se } y_1 = x_1 + 1, y_2 = x_2 \text{ e } y_3 = x_3 - 1, \\ \lambda & \text{se } y_1 = x_1 + 1, y_2 = x_2 \text{ e } y_3 = x_3, \\ 0,2\mu_2 & \text{se } y_1 = x_1, y_2 = x_2 - 1 \text{ e } y_3 = x_3, \\ 0,8\mu_3 & \text{se } y_1 = x_1, y_2 = x_2 \text{ e } y_3 = x_3 - 1, \end{cases} \quad (4.5)$$

4.3.2 Equações de trânsito - Modelo 3

As equações de trânsito (3.39) são dadas por

$$\begin{cases} \beta_1 = 0,5 + 0,2\beta_3, \\ \beta_2 = 0,5\beta_1, \\ \beta_3 = 0,5\beta_1 + 0,8\beta_2. \end{cases} \quad (4.6)$$

E as soluções dessas equações são

$$\begin{cases} \beta_1 = 0,61 \\ \beta_2 = 0,31 \\ \beta_3 = 0,55. \end{cases} \quad (4.7)$$

4.3.3 Medidas de performance - Modelo 3

Usando (3.42)-(3.46), vamos calcular as medidas de performance para esse modelo.

Para o primeiro nó, temos:

$$U = \frac{\beta_1}{\delta_1} = \frac{0,61}{1} = 0,61$$

$$Q = \frac{\beta_1}{\delta_1 - \beta_1} = \frac{0,61}{1 - 0,61} = 1,56$$

$$L = \frac{\beta_1}{\delta_1 - \beta_1} - \frac{\beta_1}{\delta_1} = \frac{0,61}{1 - 0,61} - \frac{0,61}{1} = 0,95$$

$$R = \frac{\beta_1}{(\delta_1 - \beta_1)\lambda} = \frac{0,61}{(1 - 0,61)0,5} = 3,13$$

$$W = \frac{\beta_1^2}{\delta_1(\delta_1 - \beta_1)\lambda} = \frac{0,61^2}{1(1 - 0,61)0,5} = 1,91$$

Para o segundo nó, temos:

$$U = \frac{\beta_2}{\delta_2} = \frac{0,305}{0,5} = 0,61$$

$$Q = \frac{\beta_2}{\delta_2 - \beta_2} = \frac{0,305}{0,5 - 0,305} = 1,56$$

$$L = \frac{\beta_2}{\delta_2 - \beta_2} - \frac{\beta_2}{\delta_2} = \frac{0,305}{0,5 - 0,305} - \frac{0,305}{0,5} = 0,95$$

$$R = \frac{\beta_2}{(\delta_2 - \beta_2)\lambda} = \frac{0,305}{(0,5 - 0,305)0,5} = 3,13$$

$$W = \frac{\beta_2^2}{\delta_2(\delta_2 - \beta_2)\lambda} = \frac{0,305^2}{0,5(0,5 - 0,305)0,5} = 1,91$$

E para o terceiro nó, temos:

$$U = \frac{\beta_3}{\delta_3} = \frac{0,55}{1} = 0,55$$

$$Q = \frac{\beta_3}{\delta_3 - \beta_3} = \frac{0,55}{1 - 0,55} = 1,22$$

$$L = \frac{\beta_3}{\delta_3 - \beta_3} - \frac{\beta_3}{\delta_3} = \frac{0,55}{1 - 0,55} - \frac{0,55}{1} = 0,67$$

$$R = \frac{\beta_3}{(\delta_3 - \beta_3)\lambda} = \frac{0,55}{(1 - 0,55)0,5} = 2,44$$

$$W = \frac{\beta_3^2}{\delta_3(\delta_3 - \beta_3)\lambda} = \frac{0,55^2}{1(1 - 0,55)0,5} = 1,34$$

As informações do Modelo 3 são passadas para o PDQ da seguinte maneira:

```
lambda <- 0.5
mu <- c(1, 0.5, 1)
A <- matrix(c(1, 0, -0.2, -0.5, 1, 0, -0.5, -0.8, 1), 3, 3, byrow = TRUE)
B <- c(0.5, 0, 0)
beta <- solve(A, B)

library(pdq)
Init("Jackson Network in PDQ-R")
wname <- c("Traffic")
rname <- c("No1", "No2", "No3")
CreateOpen(wname, lambda)
SetWUnit("Msg")
SetTUnit("Sec")
for (i in 1:length(rname)) {
  CreateNode(rname[i], CEN, FCFS)
  SetDemand(rname[i], wname, beta[i]/(lambda * mu[i]))
}
Solve(CANON)
Report()
```

Os resultados obtidos com o PDQ são os mesmo que foram obtidos analiticamente através da Rede de Jackson.

5 Simulação

Neste capítulo serão apresentados todos os algoritmos necessários para a simulação dos modelos apresentados e uma breve explicação de cada um.

5.1 Simulação - Modelo 1

Para a simulação do Modelo 1, serão usadas duas funções auxiliares, representadas nos Algoritmos 1 e 2.

A função `decide()` informa o próximo acontecimento do sistema, se será a chegada ou saída de um cliente, representados por 1 e -1 , respectivamente.

Algoritmo 1: Função auxiliar: `decide`

Entrada: $s1, s2$, ou seja, conteúdo do primeiro e do segundo servidor

Saída: 1 ou -1 , ou seja, chegou ou saiu alguém do sistema

se $s1 = 0$ **e** $s2 = 0$ **então**

 | $p = 1$

senão

 | **se** $s1 \neq 0$ **e** $s2 \neq 0$ **então**

 | $p = \frac{\lambda}{2\mu + \lambda}$

senão

 | $p = \frac{\lambda}{\mu + \lambda}$

$x \leftarrow \begin{cases} 1 & \text{com probabilidade } p, \\ 0 & \text{com probabilidade } 1 - p \end{cases}$

se $x = 1$ **então retorne** 1

retorne -1

A função `deslocamento()` informa o tempo para o próximo acontecimento do sistema, ou seja, quanto tempo vai demorar para chegar ou sair alguém do sistema.

Com as funções auxiliares definidas, é apresentado agora o Algoritmo 3, que trás a

Algoritmo 2: Função auxiliar: deslocamento

Entrada: $s1, s2$, ou seja, conteúdo do primeiro e do segundo servidor**Saída:** Tempo de deslocamento**se** $s1 = 0$ e $s2 = 0$ **então** | **retorne** *Exponencial*(λ)**se** $s1 \neq 0$ e $s2 \neq 0$ **então** | **retorne** *Exponencial*($2\mu + \lambda$)**retorne** *Exponencial*($\mu + \lambda$)

parte principal da simulação, e que deve ser repetida algumas vezes, já que será considerada a média das variáveis de interesse.

Ainda como parte da simulação principal, o Algoritmo 4, mostra como as medidas de performance são calculadas com as variáveis geradas pelo Algoritmo 3.

Para as medidas de performances Tempo de Espera e Tempo de Permanência no Sistema, criadas no Algoritmo 4, foi usado o Algoritmo 5, que também será utilizado nos próximos modelos.

Assim, basta calcular as médias das medidas de performances encontradas em cada simulação.

5.2 Simulação - Modelo 2

Assim como na simulação do Modelo 1, aqui também serão usadas duas funções auxiliares, representadas nos Algoritmos 6 e 7.

A função `decide()` informa o próximo acontecimento do sistema, se será a chegada ou saída de um cliente e em qual nó isso aconteceu. Para isso, a função retorna um vetor de tamanho 2, o primeiro valor do vetor representa o que aconteceu no primeiro nó e o último valor do vetor o que aconteceu no segundo nó. Logo, quando a função pode retornar:

- (1, 0) chegada de um cliente de fora do sistema para o nó 1;
- (-1, 1) passagem de um cliente do nó 1 para o nó 2;
- (0, -1) saída de um cliente do nó 2 para fora do sistema.

Algoritmo 3: Simulação da cadeia de Markov em tempo contínuo**Dados:** μ, λ **Resultado:** Algumas Medidas de Performancetamanho_da_fila $\leftarrow c(0)$; tempo_para_mudar $\leftarrow c(0)$ indice $\leftarrow 1$ n $\leftarrow 0$; serv1 $\leftarrow 0$; serv2 $\leftarrow 0$ t $\leftarrow NULL$; t[1] $\leftarrow deslocamento()$ fila_de_espera $\leftarrow list()$; fila_de_espera[[1]] $\leftarrow numeric(0)$ **enquanto** t[indice] ≤ 1000 **faça** indice $\leftarrow indice + 1$ serv1[indice] $\leftarrow serv1[indice - 1]$ serv2[indice] $\leftarrow serv2[indice - 1]$ fila_de_espera[[indice]] $\leftarrow fila_de_espera[[indice - 1]]$ tamanho_da_fila[indice] $\leftarrow tamanho_da_fila[indice - 1] + decide()$ **se** (tamanho_da_fila[indice] > tamanho_da_fila[indice - 1]) **então** n $\leftarrow n + 1$ fila_de_espera[[indice]] $\leftarrow c(fila_de_espera[[indice - 1]], n)$ **senão** **se** (serv1[indice - 1] $\neq 0$ e serv2[indice - 1] $\neq 0$) **então** **se** (rbinom(1, 1, 1/2) + 1 = 1) **então** serv1[indice] $\leftarrow 0$ **senão** serv2[indice] $\leftarrow 0$ **senão** **se** (serv2[indice - 1] = 0) **então** serv1[indice] $\leftarrow 0$ **senão** serv2[indice] $\leftarrow 0$ **se** (length(fila_de_espera[[indice]] > 0)) **então** **se** (serv1[indice] = 0 e serv2[indice] $\neq 0$) **então** serv1[indice] $\leftarrow fila_de_espera[[indice]][1]$ fila_de_espera[[indice]] $\leftarrow fila_de_espera[[indice]][-1]$ **senão** **se** (serv1[indice] $\neq 0$ e serv2[indice] = 0) **então** serv2[indice] $\leftarrow fila_de_espera[[indice]][1]$ fila_de_espera[[indice]] $\leftarrow fila_de_espera[[indice]][-1]$ **senão** **se** (serv1[indice] = 0 e serv2[indice] = 0) **então** **se** (rbinom(1, 1, 1/2) + 1 = 1) **então** serv1[indice] $\leftarrow fila_de_espera[[indice]][1]$ fila_de_espera[[indice]] $\leftarrow fila_de_espera[[indice]][-1]$ **senão** serv2[indice] $\leftarrow fila_de_espera[[indice]][1]$ fila_de_espera[[indice]] $\leftarrow fila_de_espera[[indice]][-1]$ tempo_para_mudar[indice] $\leftarrow deslocamento()$ t[indice] $\leftarrow t[indice - 1] + tempo_para_mudar[indice]$ **Retorne** Medidas de Performance

Algoritmo 4: Medidas de Performance

Dados: Simulação da cadeia de Markov em tempo contínuo**Resultado:** Medidas de Performance desejadasTransferência $\leftarrow \lambda$ Utilização $\leftarrow (sum(tempo_para_mudar[serv1 \neq 0]) + sum(tempo_para_mudar[serv2 \neq 0])) / (2 * t[indice])$ Tamanho da Fila $\leftarrow sum(tamanho_da_fila * tempo_para_mudar) / t[indice]$ tamanho_fila_de_espera $\leftarrow NULL$ **para** (k de 1 até indice) **faça** $tamanho_fila_de_espera[k] \leftarrow length(fila_de_espera[[k]])$

Tamanho da Fila de Espera

 $\leftarrow sum(tamanho_fila_de_espera * tempo_para_mudar) / t[indice]$ Tempo de Espera $\leftarrow mean(permanencia(fila_de_espera))$ lista_filas $\leftarrow list()$ **para** (m de 1 até indice) **faça** $lista_filas[[m]] \leftarrow c(serv1[m], serv2[m], fila_de_espera[[m]])$ Tempo de Permanência no Sistema $\leftarrow mean(permanencia(lista_filas))$

Algoritmo 5: Função auxiliar: permanencia

Dados: lista**Resultado:** Vetor com os tempos de permanência no sistema de todos os clientestempo_permanencia $\leftarrow NULL$ **para** k de 1 até n **faça** tempo_permanencia[k] $\leftarrow 0$ **para** j de 1 até $length(lista)$ **faça** **se** no vetor lista[j] tem pelo menos um k **então** tempo_permanencia[k] $\leftarrow tempo_permanencia[k] + tempo_para_mudar[j]$ **Retorne**(tempo_permanencia)

Algoritmo 6: Função auxiliar: decide

Entrada: $s1, s2$, ou seja, conteúdo do primeiro e do segundo servidor**Saída:** Um vetor que indica chegadas para cada nó (se negativo indica saída)

$$x \leftarrow \begin{cases} 1 & \text{com probabilidade } \frac{\lambda}{\mu_1 + \mu_2 + \lambda}, \\ 2 & \text{com probabilidade } \frac{\mu_1}{\mu_1 + \mu_2 + \lambda}, \\ 3 & \text{com probabilidade } \frac{\mu_2}{\mu_1 + \mu_2 + \lambda} \end{cases}$$

se $x = 1$ **então**| **retorne** $(1, 0)$ **senão**| **se** $x = 2$ **e** $s1 \neq 0$ **então**| | **retorne** $(-1, 1)$ | **senão**| | **se** $x = 3$ **e** $s2 \neq 0$ **então**| | | **retorne** $(0, -1)$ | | **senão**| | | **retorne** $(0, 0)$

A função `deslocamento()` informa o tempo para o próximo acontecimento do sistema. Veja o Algoritmo 7.

Algoritmo 7: Função auxiliar: deslocamento

Saída: Tempo de deslocamento**retorne** $Exponencial(\mu_1 + \mu_2 + \lambda)$

Com as funções auxiliares definidas, é apresentado agora o Algoritmo 8, que trás a parte principal da simulação, e que deve ser repetida algumas vezes, já que será considerada a média das variáveis de interesse.

Ainda como parte da simulação principal, o Algoritmo 9, mostra como as medidas de performance de cada um dos dois nós ($x = 1, 2$) são calculadas com as variáveis geradas pelo Algoritmo 8.

Para as medidas de performances Tempo de Espera e Tempo de Permanência no Sistema, criadas no Algoritmo 9, foi utilizado o Algoritmo 5, o mesmo utilizado no Modelo 1. Assim, basta calcular as médias das medidas de performances encontradas em cada simulação.

Algoritmo 8: Simulação da cadeia de Markov em tempo contínuo**Dados:** μ_1, μ_2, λ **Resultado:** Algumas Medidas de Performance

```

tamanho_da_fila ← list()
fila_de_espera_1 ← list(); fila_de_espera_2 ← list()
tamanho_da_fila[[1]] ← c(0, 0)
tempo_para_mudar ← c(0)
indice ← 1
n ← 0; serv1 ← 0; serv2 ← 0
t ← NULL; t[1] ← deslocamento()
fila_de_espera1[[1]] ← numeric(0); fila_de_espera2[[1]] ← numeric(0)

enquanto t[indice] ≤ 2000 faça
  indice ← indice + 1
  serv1[indice] ← serv1[indice - 1]
  serv2[indice] ← serv2[indice - 1]
  fila_de_espera1[[indice]] ← fila_de_espera1[[indice - 1]]
  fila_de_espera2[[indice]] ← fila_de_espera2[[indice - 1]]
  tamanho_da_fila[[indice]] ← tamanho_da_fila[[indice - 1]] + decide()
  se (tamanho_da_fila[[indice]][1] > tamanho_da_fila[[indice - 1]][1]) então
    n ← n + 1
    fila_de_espera1[[indice]] ← c(fila_de_espera1[[indice - 1]], n)
  senão
    se (tamanho_da_fila[[indice]][1] < tamanho_da_fila[[indice - 1]][1]) então
      serv1[indice] ← 0
      fila_de_espera2[[indice]] ← c(fila_de_espera2[[indice - 1]], serv1[indice - 1])
    senão
      se (tamanho_da_fila[[indice]][2] < tamanho_da_fila[[indice - 1]][2]) então
        serv2[indice] ← 0
        fila_de_espera2[[indice]] ← c(fila_de_espera2[[indice - 1]], serv1[indice - 1])

  se (length(fila_de_espera1[[indice]] > 0)) então
    se (serv1[indice] = 0) então
      serv1[indice] ← fila_de_espera1[[indice]][1]
      fila_de_espera1[[indice]] ← fila_de_espera1[[indice]][-1]
  se (length(fila_de_espera2[[indice]] > 0)) então
    se (serv2[indice] = 0) então
      serv2[indice] ← fila_de_espera2[[indice]][1]
      fila_de_espera2[[indice]] ← fila_de_espera2[[indice]][-1]

  tempo_para_mudar[indice] ← deslocamento()
  t[indice] ← t[indice - 1] + tempo_para_mudar[indice]

```

Retorne Medidas de Performance

Algoritmo 9: Medidas de Performance

Dados: Simulação da cadeia de Markov em tempo contínuo

Resultado: Medidas de Performance desejadas para o nó "x"

Transferência_x $\leftarrow \lambda$

Utilização_x $\leftarrow (\text{sum}(\text{tempo_para_mudar}[\text{servx} \neq 0])) / (t[\text{indice}])$

tamanho_da_filax $\leftarrow \text{NULL}$

para (k de 1 até indice) **faça**

 | tamanho_da_filax[k] \leftarrow tamanho_da_filax[[k]][x]

Tamanho da Fila_x $\leftarrow \text{sum}(\text{tamanho_da_filax} * \text{tempo_para_mudar}) / t[\text{indice}]$

tamanho_fila_de_esperax $\leftarrow \text{NULL}$

para (k de 1 até indice) **faça**

 | tamanho_fila_de_esperax[k] $\leftarrow \text{length}(\text{fila_de_esperax}[[k]])$

Tamanho da Fila de Espera_x

$\leftarrow \text{sum}(\text{tamanho_fila_de_esperax} * \text{tempo_para_mudar}) / t[\text{indice}]$

Tempo de Espera_x $\leftarrow \text{mean}(\text{permanencia}(\text{fila_de_esperax}))$

listax_filas $\leftarrow \text{list}()$

para (m de 1 até indice) **faça**

 | listax_filas[[m]] $\leftarrow c(\text{servx}[m], \text{fila_de_esperax}[[m]])$

Tempo de Permanência no Sistema_x $\leftarrow \text{mean}(\text{permanencia}(\text{listax_filas}))$

5.3 Simulação - Modelo 3

Para a simulação do Modelo 3 também serão usadas duas funções auxiliares, representadas nos Algoritmos 10 e 11.

A função `decide()` informa o próximo acontecimento do sistema, se será a chegada ou saída de um cliente e em qual nó isso aconteceu. Para isso, a função retorna um vetor de tamanho 3, já que o modelo possui 3 nós.

Algoritmo 10: Função auxiliar: `decide`

Entrada: $s1, s2, s3$, ou seja, conteúdo dos três servidores

Saída: vetor que representa a chegada ou saída de clientes em cada nó

$$d = 0.5\mu_1 + 0.5\mu_1 + 0.8\mu_2 + 0.2\mu_3 + \lambda + 0.2\mu_2 + 0.8\mu_3$$

$$x \leftarrow \begin{cases} 1 & \text{com probabilidade } \frac{0.5\mu_1}{d}, \\ 2 & \text{com probabilidade } \frac{0.5\mu_1}{d}, \\ 3 & \text{com probabilidade } \frac{0.8\mu_2}{d}, \\ 4 & \text{com probabilidade } \frac{0.2\mu_3}{d}, \\ 5 & \text{com probabilidade } \frac{\lambda}{d}, \\ 6 & \text{com probabilidade } \frac{0.2\mu_2}{d}, \\ 7 & \text{com probabilidade } \frac{0.8\mu_3}{d} \end{cases}$$

se $x = 1$ **e** $s1 \neq 0$ **então**

retorne $(-1, 1, 0)$

se $x = 2$ **e** $s1 \neq 0$ **então**

retorne $(-1, 0, 1)$

se $x = 3$ **e** $s2 \neq 0$ **então**

retorne $(0, -1, 1)$

se $x = 4$ **e** $s3 \neq 0$ **então**

retorne $(1, 0, -1)$

se $x = 5$ **então**

retorne $(1, 0, 0)$

se $x = 6$ **e** $s2 \neq 0$ **então**

retorne $(0, -1, 0)$

se $x = 7$ **e** $s3 \neq 0$ **então**

retorne $(0, 0, -1)$

retorne $(0, 0)$

A função `deslocamento()` informa o tempo para o próximo acontecimento do sistema.

Veja o Algoritmo 11.

Algoritmo 11: Função auxiliar: deslocamento

Saída: Tempo de deslocamento

$$d = 0.5\mu_1 + 0.5\mu_1 + 0.8\mu_2 + 0.2\mu_3 + \lambda + 0.2\mu_2 + 0.8\mu_3$$

retorne *Exponencial*(d)

Com as funções auxiliares definidas, é apresentado agora o Algoritmo 12, que trás a parte principal da simulação, e que deve ser repetida algumas vezes, já que será considerada a média das variáveis de interesse.

Para calcular as medidas de performance, pode ser utilizado o Algoritmo 9, o mesmo do Modelo 2, sendo que agora $x = 1, 2, 3$. O Algoritmo 5, o mesmo para todos os modelos, será utilizado nas medidas de performances Tempo de Espera e Tempo de Permanência no Sistema. Assim, basta calcular as médias das medidas de performances encontradas em cada simulação.

Algoritmo 12: Simulação da cadeia de Markov em tempo contínuo**Dados:** μ (vetor de tamanho 3), λ **Resultado:** Algumas Medidas de Performance

```

tamanho_da_fila  $\leftarrow$  list(); fila_de_espera_1  $\leftarrow$  list()
fila_de_espera_2  $\leftarrow$  list(); fila_de_espera_3  $\leftarrow$  list()
tamanho_da_fila[[1]]  $\leftarrow$  c(0,0,0); tempo_para_mudar  $\leftarrow$  c(0)
indice  $\leftarrow$  1; n  $\leftarrow$  0; serv1  $\leftarrow$  0; serv2  $\leftarrow$  0; serv3  $\leftarrow$  0
t  $\leftarrow$  NULL; t[1]  $\leftarrow$  deslocamento(); fila_de_espera1[[1]]  $\leftarrow$  numeric(0)
fila_de_espera2[[1]]  $\leftarrow$  numeric(0); fila_de_espera3[[1]]  $\leftarrow$  numeric(0)

```

enquanto $t[\text{indice}] \leq 6000$ **faça**

```

    indice  $\leftarrow$  indice + 1
    serv1[indice]  $\leftarrow$  serv1[indice - 1]
    serv2[indice]  $\leftarrow$  serv2[indice - 1]
    serv3[indice]  $\leftarrow$  serv3[indice - 1]
    fila_de_espera1[[indice]]  $\leftarrow$  fila_de_espera1[[indice - 1]]
    fila_de_espera2[[indice]]  $\leftarrow$  fila_de_espera2[[indice - 1]]
    fila_de_espera3[[indice]]  $\leftarrow$  fila_de_espera3[[indice - 1]]
    tamanho_da_fila[[indice]]  $\leftarrow$  tamanho_da_fila[[indice - 1]] + decide()
    se (sum(tamanho_da_fila[[indice - 1]] + c(1,0,0) = tamanho_da_fila[[indice]]) = 3) então
        n  $\leftarrow$  n + 1
        fila_de_espera1[[indice]]  $\leftarrow$  c(fila_de_espera1[[indice - 1]], n)
    se (sum(tamanho_da_fila[[indice - 1]] + c(-1,0,0) = tamanho_da_fila[[indice]]) = 3) então
        serv1[indice]  $\leftarrow$  0
    se (sum(tamanho_da_fila[[indice - 1]] + c(-1,0,1) = tamanho_da_fila[[indice]]) = 3) então
        serv1[indice]  $\leftarrow$  0
        fila_de_espera3[[indice]]  $\leftarrow$  c(fila_de_espera3[[indice - 1]], serv1[indice - 1])
    se (sum(tamanho_da_fila[[indice - 1]] + c(0,-1,0) = tamanho_da_fila[[indice]]) = 3) então
        serv2[indice]  $\leftarrow$  0
        fila_de_espera3[[indice]]  $\leftarrow$  c(fila_de_espera3[[indice - 1]], serv2[indice - 1])
    se (sum(tamanho_da_fila[[indice - 1]] + c(0,0,-1) = tamanho_da_fila[[indice]]) = 3) então
        serv3[indice]  $\leftarrow$  0
    se (sum(tamanho_da_fila[[indice - 1]] + c(1,0,-1) = tamanho_da_fila[[indice]]) = 3) então
        serv3[indice]  $\leftarrow$  0
        fila_de_espera1[[indice]]  $\leftarrow$  c(fila_de_espera1[[indice - 1]], serv3[indice - 1])
    se (length(fila_de_espera1[[indice]] > 0) então
        se (serv1[indice] = 0) então
            serv1[indice]  $\leftarrow$  fila_de_espera1[[indice]][1]
            fila_de_espera1[[indice]]  $\leftarrow$  fila_de_espera1[[indice]][-1]
    se (length(fila_de_espera2[[indice]] > 0) então
        se (serv2[indice] = 0) então
            serv2[indice]  $\leftarrow$  fila_de_espera2[[indice]][1]
            fila_de_espera2[[indice]]  $\leftarrow$  fila_de_espera2[[indice]][-1]
    se (length(fila_de_espera3[[indice]] > 0) então
        se (serv3[indice] = 0) então
            serv3[indice]  $\leftarrow$  fila_de_espera3[[indice]][1]
            fila_de_espera3[[indice]]  $\leftarrow$  fila_de_espera3[[indice]][-1]
    tempo_para_mudar[indice]  $\leftarrow$  deslocamento()
    t[indice]  $\leftarrow$  t[indice - 1] + tempo_para_mudar[indice]

```

Retorne Medidas de Performance

6 Resultados

Neste capítulo serão apresentados os resultados para os três modelos abordados neste trabalho obtidos através da simulação via Cadeias de Markov e do PDQ.

6.1 Resultados - Modelo 1

Para chegarmos ao resultado final das medidas de performance desejadas, a simulação foi repetida 50 vezes com o tempo $t = 1000$ em cada vez, assim já foi possível encontrar resultados satisfatórios se comparados aos obtidos através do PDQ, como pode ser visto na Tabela 1.

Tabela 1: Resultados do Modelo 1

	PDQ	Simulação
Taxa de Transferência	0,75	0,75
Utilização (porcentagem)	37,50	37,63
Tamanho da Fila	0,87	0,87
Tamanho da Fila de Espera	0,12	0,12
Tempo de Espera	0,16	0,15
Tempo de Permanência no Sistema	1,16	1,15

Em um sistema com dois servidores e uma única fila, onde os momentos de chegadas dos clientes ao sistema formam um processo de Poisson Homogêneo com taxa $\lambda = 0,75$ e o tempo de serviço de cada servidor é distribuído exponencialmente com taxa $\mu = 1$, é esperado uma média de utilização dos servidores de 37,5%.

O tamanho médio da fila, considerando quem está sendo servido, é 0,87 clientes. Enquanto o tamanho médio da fila de espera será de apenas 0,12 clientes.

O tempo médio gasto esperando para ser atendido é de 0,16 minuto já o tempo médio gasto em todo o processo, ou seja, o tempo gasto esperando e sendo atendido é de 1,16 minuto.

6.2 Resultados - Modelo 2

Para chegarmos ao resultado final das medidas de performance desejadas para esse modelo, a simulação foi repetida 50 vezes com o tempo $t = 2000$ em cada vez, assim já foi possível encontrar resultados satisfatórios se comparados aos obtidos através do PDQ, como pode ser visto na Tabela 2, que apresenta essas medidas para os dois nós separadamente.

Tabela 2: Resultados do Modelo 2

	PDQ 1	Simulação 1	PDQ 2	Simulação 2
Taxa de Transferência	0,75	0,75	0,75	0,75
Utilização (porcentagem)	18,75	18,60	56,25	55,64
Tamanho da Fila	0,23	0,22	1,29	1,20
Tamanho da Fila de Espera	0,04	0,04	0,72	0,64
Tempo de Espera	0,06	0,05	0,96	0,87
Tempo de Permanência no Sistema	0,31	0,30	1,71	1,62

Em um sistema com dois servidores em série, onde os momentos de chegadas dos clientes ao sistema formam um processo de Poisson Homogêneo com taxa $\lambda = 0,75$ e o tempo de serviço de cada servidor é distribuído exponencialmente com taxa $\mu_1 = 4$ e $\mu_2 = 1,33$ respectivamente, é esperado uma média de utilização de aproximadamente 18,7% no primeiro servidor e 56,2% no segundo.

O tamanho médio da fila, considerando quem está sendo servido, é 0,2 cliente no primeiro nó e 1,2 cliente no segundo. Enquanto o tamanho médio da fila de espera é 0,04 cliente no primeiro e 0,7 cliente no segundo.

O tempo médio gasto esperando para ser atendido é de 0,06 minuto no primeiro nó e 0,9 no segundo, já o tempo médio gasto em todo o processo, ou seja, o tempo gasto esperando e sendo atendido é de 0,3 minuto no primeiro nó e 1,7 minutos no segundo.

6.3 Resultados - Modelo 3

Para chegarmos ao resultado final das medidas de performance desejadas para esse modelo, a simulação foi repetida 50 vezes com o tempo $t = 6000$ em cada vez, assim já foi possível encontrar resultados satisfatórios se comparados aos obtidos através do PDQ, como pode ser visto na Tabela 3, que apresenta essas medidas para os três nós separadamente.

Tabela 3: Resultados do Modelo 3

	PDQ 1	S 1	PDQ 2	S 2	PDQ 3	S 3
Taxa de Transferência	0,50	0,50	0,50	0,50	0,50	0,50
Utilização (porcentagem)	60,98	59,84	60,98	59,47	54,88	53,55
Tamanho da Fila	1,56	1,47	1,56	1,56	1,22	1,15
Tamanho da Fila de Espera	0,95	0,88	0,95	0,97	0,67	0,62
Tempo de Espera	1,91	1,80	1,91	1,99	1,33	1,27
Tempo de Permanência no Sistema	3,12	3,03	3,12	3,21	2,43	2,37

Em um sistema de rede de comunicações como o modelo 3, é esperado uma média de utilização de aproximadamente 60%, 60% e 54% em cada servidor, respectivamente.

O tamanho médio da fila, considerando quem está sendo servido, é 1,5 para os dois primeiros nós e 1,2 no terceiro nó. Enquanto o tamanho médio da fila de espera é 0,9 nos dois primeiros e 0,6 no terceiro.

O tempo médio gasto esperando para ser atendido em cada servidor é 1,9, 1,9 e 1,3 minutos, respectivamente. Já o tempo médio gasto em todo o processo, ou seja, o tempo gasto esperando e sendo atendido é de 3,1 minutos nos dois primeiros nós e 2,4 minutos no terceiro nó.

7 Conclusões

Foram estudadas duas classes de métodos que podem ser aplicadas para a análise de problemas ligados aos sistemas de filas, métodos analíticos através do pacote PDQ no R e métodos de simulação através da simulação via Cadeias de Markov. Foram apresentados três modelos onde foi possível resolvê-los das duas maneiras.

Foi mostrado como chegar às medidas de performance que o PDQ calcula para cada modelo. No modelo 1, sistema de um nó com servidor duplo, esses cálculos analíticos foram realizados através de processos de nascimento e morte. E para os modelos 2 e 3, dois servidores em série e rede de comunicações, respectivamente, os cálculos analíticos foram realizados através de Redes de Jackson.

No modelo 1, com pouco tempo de simulação já foi possível chegar aos resultados encontrados analiticamente. Já nos outros modelos, o tempo de simulação teve que ser aumentado para observarmos melhores resultados. É possível concluir que os dois métodos são eficientes e apresentaram resultados bastante semelhantes.

Referências

- [1] GUNTHER, N. J. *Analyzing Computer System Performance with Perl::PDQ (2. ed.)*. [S.l.]: Springer, 2011. I-XXVIII, 1-474 p. ISBN 978-3-642-22582-6.
- [2] GUNTHER, N. J. *The practical performance analyst : performance-by-design techniques for distributed systems*. New York: MCGRAW-HILL, 1998. (McGraw-Hill series on computer communications). ISBN 0-07-912946-3.
- [3] ROSS, S. M. *Introduction to probability models*. Ninth. [S.l.]: Harcourt/Academic Press, Burlington, MA, 2007.
- [4] ASMUSSEN, S. *Applied probability and queues*. Second. [S.l.]: Springer-Verlag, New York, 2003. xii+438 p. (Applications of Mathematics (New York), v. 51). Stochastic Modelling and Applied Probability. ISBN 0-387-00211-1.
- [5] ANDRADE, E. de. *Introdução à pesquisa operacional: métodos e modelos para análise de decisões*. LTC, 1998. ISBN 9788521614128. Disponível em: <<http://books.google.com.br/books?id=bUmLGAAACAAJ>>.
- [6] FERRARI, P. A.; GALVES, A. *Acoplamento em processos estocásticos*. First. [s.n.], 1997. Hardcover. Disponível em: <<http://www.ime.usp.br/~pablo/papers/libro.pdf>>.